Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

E. JAMES ARCHER, *University of Wisconsin*
JUDSON S. BROWN, *State University of Iowa*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *Ohio State University*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*

CONTENTS

Marginal and Conditional Stimulus and Response Probabilities in Verbal Conditioning: J. ENGLER 303
Long-Term Retention of Perceptual-Motor Skills: R. B. AMMONS, R. G. FARR, E. BLOCH, E. NEUMANN, M. DEY, R. MARION, AND C. H. AMMONS 318
Meaning, Frequency, and Visual Duration Threshold: J. A. TAYLOR 329
Acquisition and Extinction of the Conditioned Eyelid Response Following Partial and Continuous Reinforcement: W. F. REYNOLDS 335
An Extension of Hullian Theory to Response Decrements Resulting From Drugs: R. B. PAYNE 342
Conditioned-Stimulus Variables in Avoidance Learning: M. SCHWARTZ 347
Age Differences in Short-Term Retention of Rapidly Changing Information: W. K. KIRCHNER 352
A New Technique for Studying Spatial Generalization With Voluntary Responses: J. S. BROWN, F. R. CLARKE, AND L. STEIN 359
The Effects of Delay of Reinforcement Upon an Operant Discrimination in the Pigeon: J. L. MYERS 363
An Analysis of Two-Person Game Situations in Terms of Statistical Learning Theory: R. C. ATKINSON AND P. SUPPES 369
Variable Frequency of Knowledge of Results and the Learning of a Simple Skill: E. A. BILODEAU AND I. MCD. BILODEAU 379
Information Processing and Problem Solving: C. M. SOLLEY AND F. W. SNYDER 384
Distance Constancy: O. W. SMITH 388
Apparent Backward Association: A Situational Effect: L. H. STORMS 390
Erratum 396

American Psychological Association
Vol. 55 No. 4 April 1958

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Circulation Manager
SADIE J. DOYLE, Editorial Assistant

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this JOURNAL, 1957, **53**, 1–2).

Address all articles submitted for publication to the Editor: Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association.*** When in doubt about practices in this Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1958 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1958 by the American Psychological Association, Inc.

EDITORIAL ANNOUNCEMENT

Professor Delos D. Wickens, who has served as Associate Editor of the JOURNAL during the processing of the manuscripts received in 1956 and 1957 for publication in 1957 and 1958, has resigned that position effective about April 1, 1958. Fortunately, he has consented to continue in the service of the JOURNAL in the role of Consulting Editor. We are pleased to announce that Professor William K. Estes has accepted appointment as Associate Editor of the JOURNAL, effective April 1, 1958.

Journal of
Experimental Psychology

VOL. 55, NO. 4 APRIL, 1958

MARGINAL AND CONDITIONAL STIMULUS AND RESPONSE PROBABILITIES IN VERBAL CONDITIONING[1]

JEAN ENGLER[2]

Northwestern University

The experiments reported in the present paper were designed to test predictions from the Bush-Mosteller stochastic model for learning (**2**) and to provide information concerning the dependence of the learning rate parameter, α, on various experimental conditions.

The experimental procedure involved the presentation of series of independent and Markov two-choice events in a Humphreys-type verbal conditioning situation. In this procedure S is given a signal indicating that a trial has begun and he then predicts which of two lights will come on. Prediction of Light 1, Response A_1, occurs with probability p; prediction of Light 2, Response A_2, occurs with probability $q = 1 - p$. After S has made his response, E lights either Light 1 (Event E_1 occurring with probability $\pi(1)$) or Light 2 (Event E_2 occurring with probability $\pi(2) = 1 - \pi(1)$) according to a prearranged schedule independent of S's response. The present paper is particularly concerned with cases in which E specifies not only the marginal probabilities, but also the conditional probabilities of Events E_1 and E_2, thus making the experimental series one of Markov, not independent, events. For such cases the probability of occurrence of Event E_1 on Trial n given (i.e., conditional on) the occurrence of Event E_2 on Trial $n - 1$ is denoted by $\pi(1|2)$. $\pi(1|1)$, $\pi(2|1)$, and $\pi(2|2)$ are similarly defined. An independent series is thus the special case in which $\pi(1|1) = \pi(1|2) = \pi(1)$ and $\pi(2|1) = \pi(2|2) = \pi(2)$.

In the terminology of the model proposed by Bush and Mosteller, the events being considered are experimenter-controlled. Furthermore, the events—the appearance of one light or the other—have intrinsic symmetry. Using this symmetry and the assumption that the limit points of their operators should be such that, if E_1 occurs on all trials, Response A_1 will occur with a probability which tends to unity, Bush and Mosteller derive the following equation:

$$V_{1,n} = \alpha^n V_{1,0} + \pi(1)(1 - \alpha^n), \quad (1)$$

where $V_{1,n}$ is the expected value of the probability, p, of occurrence of Response A_1 on Trial n; $V_{1,0}$ is the expected value of the probability of occurrence of Response A_1 on the first trial (estimated from the data or assumed a priori);

[1] Based on a dissertation submitted to Northwestern University while the author was a predoctoral fellow of National Science Foundation. Thanks are expressed to John W. Cotton who directed the research.

[2] Now at Institute of Statistics, University of North Carolina, Chapel Hill, N. C.

$\pi(1)$ is the probability of occurrence of Event E_1; and α is the learning rate parameter estimated from the data.

Equation 1 is valid for series of either independent or Markov events. In the Markov case, however, the assumptions used in obtaining Equation 1 imply also

$$V_{1,n}|E_1 = \alpha^n[V_{1,0} - \pi(1)][1 - K'^{n-1}] + (K'\alpha)^{n-1}[\alpha(V_{1,0} - 1) + 1] + [K\pi(1)\alpha + (1 - \alpha)]\left[\frac{1 - (K'\alpha)^{n-1}}{1 - K'\alpha}\right] \quad (2a)$$

$$V_{1,n}|E_2 = \alpha^n[V_{1,0} - \pi(1)][1 - K'^{n-1}] + (K'\alpha)^{n-1}V_{1,0} + K\pi(1)\alpha\left[\frac{1 - (K'\alpha)^{n-1}}{1 - K'\alpha}\right], \quad (2b)$$

where $K = \pi(1|2)/\pi(1)$ and $K' = \pi(1|1) - \pi(1|2)$. $V_{1,n}|E_1$ is the expected value of the probability of occurrence of Response A_1 on Trial n, given Event E_1 on Trial $n - 1$; $V_{1,n}|E_2$ is the expected value of the probability of occurrence of Response A_1 on Trial n, given Event E_2 on Trial $n - 1$; and $V_{1,0}$, α, $\pi(1|1)$, $\pi(1/2)$, $\pi(2|1)$, and $\pi(2|2)$ are defined as above. Note that the ordinate values on the graph of the empirical curve corresponding to Equation 2a are the proportions of *S*s who make Response A_1 on a trial immediately preceded by a trial on which Event E_1 occurred. Only the proportion $\pi(1)$ of the data, then, will be available for plotting this curve. Similarly, the proportion $\pi(2)$ of the data will be available for plotting the curve corresponding to Equation 2b. It thus becomes imperative that different series (having the same probability specifications) be used for different *S*s within a group if a value is to be obtained for each trial.[3]

A number of experiments (**4, 5, 6, 7, 8, 9**) indicate that there is, in general, close agreement between the probability of an experimental event and the asymptotic relative frequency of its prediction by *S*s, at least for two-choice situations. Equation 1 predicts this result and also, quite accurately, the shape of the learning curves.

In the only reported study providing a test of the accuracy of Equation 2, Hake and Hyman (**7**) investigated both independent and Markov series of experimental events using independent groups of *S*s for each series. Re-analysis of their data[4] indicated that asymptotic conditional response probabilities closely approximated conditional stimulus probabilities. This result is a contradiction of Equation 2 whose asymptote is a function of α and, for Case (a), varies between $\pi(1)$ and 1. A more detailed comparison of the obtained results with those predicted by the model was not possible since every *S* in a given group had the same series and the data are thus unsuitable for parameter estimation.

The experiments of the present study were therefore designed to provide data for a more detailed test of the accuracy of Equation 2. The results provide empirical evidence concerning the relationship of asymptotic marginal and conditional response probabilities and the corresponding learning curves and learning rates to (*a*) conditional probability specifications; (*b*) marginal probability specifications; (*c*) conditional probability specifications for fixed marginal probability specifications; (*d*) prior experience with various types of experimental series; and (*e*) difference between initial and asymptotic response behavior.

[3] Even though trials are eventually combined in graphing the curve, this procedure is necessary for valid parameter estimation.

[4] Thanks are expressed to Dr. Hake who promptly granted a request for the data so that this analysis, not relevent to his own considerations of information theory, could be made.

METHOD

Common Procedure

Apparatus.—A brown beaver-board screen 42 × 34 in. separated the apparatus visible to *S* from the apparatus visible to *E*. On *S*'s side of the screen were two brown boxes 4 × 4 × 2 in. On each box a red light 1 in. in diameter and a toggle switch were mounted on the vertical side facing *S*. From these boxes wires ran under the screen to the apparatus on *E*'s side of the screen. On this side were mounted a toggle switch which started a buzzer, two 15-w. bulbs each of which was lighted by the toggle switch on the appropriate box on *S*'s side of the screen, and two toggle switches each of which lighted the red light on the appropriate box on *S*'s side of the screen. The toggle switch on each of *S*'s boxes turned off the buzzer as well as turning on the appropriate light on *E*'s side of the screen. The two boxes in front of *S* and the corresponding lights were arbitrarily labeled E_1 and E_2 as a recording device for *E*. The *S* was not aware of this labeling.

Subjects.—The *S*s were 280 students from undergraduate psychology courses at Northwestern University. Eighty *S*s were assigned to four random experimental groups of 20 *S*s each for Exp. I and 200 *S*s were assigned to 10 random experimental groups of 20 *S*s each for Exp. II. None of the *S*s had previously served in a prediction experiment.

Procedure.—Each *S* in Exp. I was given 480 consecutive trials and each *S* in Exp. II was given 240 consecutive trials. The trials were run off without a break. The *E* signaled the beginning of each trial by sounding a buzzer, *S* made his choice by pressing the toggle switch on the appropriate box, and *E* indicated the correct choice to *S* by lighting the light on the corresponding box. A timer was wired into the apparatus and *E* sounded the buzzer beginning a trial every 5 sec.

Instructions.—At the beginning of the experimental session *S* was brought into the room, seated, and given the following instructions:

"As you can see, there is very little apparatus on your side of the screen—only these two brown boxes. You will notice that each box has a red light and a choice button.

"There will be a long series of trials. For each trial one of the boxes has been arbitrarily chosen as correct. Your task will be to guess which box is coming up next each time. You can indicate your choice by pushing one of the choice buttons. Please keep one hand ready to push the button on each box during the experiment. I will indicate the beginning of a trial by sounding this buzzer [*E* sounded buzzer and turned it off]. When you push one of the choice buttons the buzzer will stop and I will know which box you have chosen. Let's try it [*E* sounded buzzer and *S* indicated his choice by pushing one of the choice buttons]. After you have made your choice, the light in the correct box will be lit, like this [*E* lighted light in box chosen by *S*]. If neither box lights, please tell me.

"Try to get as many right as possible. Each time choose as quickly as you can, after the buzzer sounds, the box which you feel is most likely to come up next and push the choice button immediately. Any questions?"

Questions were answered by repeating or paraphrasing the instructions.

TABLE 1

MARGINAL AND CONDITIONAL PROBABILITY SPECIFICATIONS DIFFERENTIATING THE EXPERIMENTAL SERIES OF EXP. I

	Series r	Series d	Series r′	Series d′
$\pi(E_1)$	.75	.75	.25	.25
$\pi(E_1 \mid E_1)$	.75	.90	.25	.70
$\pi(E_2)$	.25	.25	.75	.75
$\pi E_2 \mid E_2)$	.25	.70	.75	.90

Experiment I

General design.—Ten different series of 120 stimuli each were constructed to comply with each of the four sets of marginal and conditional probability specifications given in Table 1.

Each *S* was given 480 trials with each successive block of 120 trials conforming to one of the sets of probability specifications. The orders in which these series were arranged differentiated the four experimental groups as shown in Table 2. Note that for each group

TABLE 2

ORDERS IN WHICH THE EXPERIMENTAL SERIES WERE ARRANGED TO FORM EXPERIMENTAL SEQUENCES FOR THE GROUPS OF EXP. I

	Trials			
	1–120	120–240	240–360	360–480
Group A	r	d	r′	d′
Group B	d	r	d′	r′
Group C	r′	d′	d	r
Group D	d′	r′	r	d

the marginal probability specifications change only after 240 trials, although the conditional probability specifications change after every 120 trials.

Stimulus materials.—The series of 120 stimuli were constructed in blocks of 24 stimuli each. Marginal probability specifications were met exactly for each of these blocks and conditional probability specifications were met as exactly as was possible. When conditional probability specifications gave fractional values for blocks of 24 trials, these specifications were approximated over series in a randomly determined fashion.

In arranging the blocks of 24 trials into series, the number of occurrences of E_1 and $E_1|E_1$ were kept constant over trials for each series insofar as was possible. The blocks were arranged so that runs did not exceed 10 consecutive occurrences for Series r and r′ and did not exceed 18 consecutive occurrences for Series d and d′.

Each of the 40 series of 120 stimuli was used once for each of the four experimental groups. There were in all, then, 40 series of 480 stimuli, 10 belonging to each of the four experimental groups. Each of these series was presented to one *S* with the box arbitrarily labeled E_1 on the right-hand side and to one *S* with the box arbitrarily labeled E_2 on the right-hand side.

Assignment of groups.—When *S*s reported to the laboratory they were assigned to the series in a randomly predetermined latin-square order more fully described elsewhere (**3**).

Experiment II

General design.—Ten different series of 240 stimuli each were constructed to comply with each of the 10 sets of marginal and conditional probability specifications given in Table 3. Note that this design may be regarded as a 2×5 factorial with two variations in marginal probability specifications and five variations in the conditional probability specification $\pi(E_1|E_1)$.

Stimulus materials.—The series of 240 stimuli were constructed in blocks of 24 stimuli with the exception of Series E and J which, because of the restrictive conditional probability specifications, were constructed in blocks of 48 stimuli. Marginal probability specifications were met exactly for each of these blocks and conditional probability specifications were met as exactly as was possible.

If each block of 24 (48) trials is considered as a position for Series A, B, C, D, F, G, and H (E and J), each of these series of 240 trials has a total of 10 (5) positions. The blocks of 24 (48) trials for a given series were constructed so that the number of occurrences of E_1 and of $E_1|E_1$ were kept approximately constant over trials. Each block occurred equally often in each position and conditional probability specifications were considered in arranging the blocks into series, so this method of constructing the blocks provided for an approximately constant number of occurrences of E_1 and $E_1|E_1$ over the trials of each series.

TABLE 3

MARGINAL AND CONDITIONAL PROBABILITY SPECIFICATIONS DIFFERENTIATING THE EXPERIMENTAL SERIES OF EXP. II

Group	$\pi(E_1)$	$\pi(E_1\|E_1)$	$\pi(E_2)$	$\pi(E_2\|E_2)$
A	.25	.10	.75	.70
B	.25	.25	.75	.75
C	.25	.50	.75	.83⅓
D	.25	.70	.75	.90
E	.25	.90	.75	.96⅔
F	.50	.10	.50	.10
G	.50	.25	.50	.25
H	.50	.50	.50	.50
I	.50	.70	.50	.70
J	.50	.90	.50	.90

There were thus 100 series of 240 stimuli, 10 belonging to each of the 10 experimental groups. Each of these series was presented to one *S* with the box arbitrarily labeled E_1 on the right-hand side and to one *S* with the box arbitrarily labeled E_2 on the right-hand side.

Assignment of groups.—When *S*s reported to the laboratory they were assigned to the series in a randomly predetermined latin-square order more fully described elsewhere (**3**).

Treatment of Results[5]

Marginal data.—The expected value of the probability of occurrence of Response A_1 on Trial n is given by Equation 1. The learning rate parameter α is estimated from the data according to the following equation:

$$\alpha = \frac{\sum_{n=0}^{t-1} V_{1,n} - (t-1)\pi(1) - V_{1,0}}{\sum_{n=0}^{t-1} V_{1,n} - t\pi(1)}, \qquad (3)$$

where t is the number of trials on which the estimate is based. Since the experimental series were constructed in blocks of 24 (or 48) trials, predictions were obtained for blocks of 24 (or 48) trials by use of the following equation:

$$\sum_{n=t'}^{t'+t} V_{1,n} = (V_{1,0} - \pi(1))\frac{\alpha^{t'} - \alpha^{t'+t}}{1-\alpha} + t\pi(1), \qquad (4)$$

[5] Derivations of all equations presented and proofs of all statements made concerning their properties are available elsewhere (**3**).

where the prediction is for the block of trials t' to $t'+t$, values being then converted to probabilities. Note that the predicted curves represented in this way are a collapsed, not a cumulative, form of Equation 1. None of the essential characteristics of the equation are obscured by representing it in this manner.

Conditional data.—The expected value of the probability of occurrence of Response A_1 on Trial n given event E_1 on Trial $n-1$ is given by Equation 2a. Note that the asymptotic term is $[K\pi(1)\alpha+(1-\alpha)][1-(K'\alpha)^{n-1}/1-K'\alpha]$ giving an asymptote of $[K\pi(1)\alpha+(1-\alpha)/1-K'\alpha]$. Since $o \leq \alpha \leq 1$ and it can be shown that the asymptote is a monotonic decreasing function of α, the asymptote varies between 1 and $\pi(1)$ as a function of α. Therefore, it is not possible for the equation to predict an asymptotic conditional response probability below the actual marginal stimulus probability of the series. Note that in Equation 1 α determines only rate of approach of marginal data to the asymptote, whereas Equation 2a implies that in conditional data the asymptote as well as the rate depends upon α. Predictions for blocks of 24 (or 48) trials were obtained by use of the following equation:

$$\sum_{n=t'}^{t'+t} V_{1,n}|E_1 = \frac{\alpha(\alpha^{t'}-\alpha^{t'+t})}{1-\alpha}[V_{1,0}-\pi(1)] + \frac{t}{1-K'\alpha}[K\pi(1)\alpha+(1-\alpha)] + \left\{\alpha[\pi(1)-1]+1-\frac{K\pi(1)\alpha+(1-\alpha)}{1-K'\alpha}\right\}\frac{(K'\alpha)^{t'}-(K'\alpha)^{t'+t}}{1-K'\alpha}, \quad (5)$$

where again prediction is for the block of trials t' to $t'+t$, values then being converted to probabilities. In the case of conditional probabilities the collapsed curves do obscure one characteristic of the trial-by-trial curves, since certain of these curves are nonmonotonic in the early trials. This oscillatory behavior is generally confined to the first 10 trials.

Results

Experiment I

Marginal data.—Predicted and obtained response probabilities for blocks of 24 trials are shown in Fig. 1. As in previous research, theory seems able to predict marginal learning curves quite accurately with respect to both shape and asymptote, regardless of conditional probability specification of the series. Dependence of marginal learning curves on conditional stimulus probability is analyzed in more detail below.

TABLE 4

Estimated Values of α Based on Marginal Data for Exp. I

Exp. Group	First 120 Trials		First 240 Trials		Second 240 Trials	
	Series	α	Series	α	Series	α
A	r	.985	r & d	.989	r′ & d′	.936
B	d	.967	d & r	.968	d′ & r′	.940
C	r′	.977	r′ & d′	.981	d & r	.905
D	d′	.964	d′ & r′	.962	r & d	.808

Estimated values of α used in plotting theoretical curves are given in Table 4. Marginal probability specifications changed after 240 trials, so estimates are based on data for 20 *Ss* for 240 trials. $V_{1,0}$ was taken as .50 for the first 240 trials for each group[6] and the predicted asymptote ($\pi(1)$) for the first 240 trials was taken as $V_{1,0}$ for the second 240 trials.[7]

Equation 1 depends only on $V_{1,0}$, α, and $\pi(1)$; its asymptote is $\pi(1)$, independent of $V_{1,0}$ and α. Therefore, the differences between the predicted learning curves for Groups A and B and for Groups C and D are entirely attributable to differences in estimated α values.

[6] .50 is, of course, the value assigned to $V_{1,0}$ a priori or on the basis of the principle of ignorance. In no case was the empirical value significantly different from .50, so use of this value for all groups seemed justified for purposes of comparability.

[7] In all cases except Group A, the predicted response probability actually reached its asymptote in the first block of 240 trials. For this group the predicted response probability was .732 at the end of 240 trials, the asymptote being .750. Nevertheless, the asymptotic value was used in this case for comparability with the other groups. Had the predicted value .732 been used as $V_{1,0}$, α for Group A would have been .938 instead of the .926 obtained using .750.

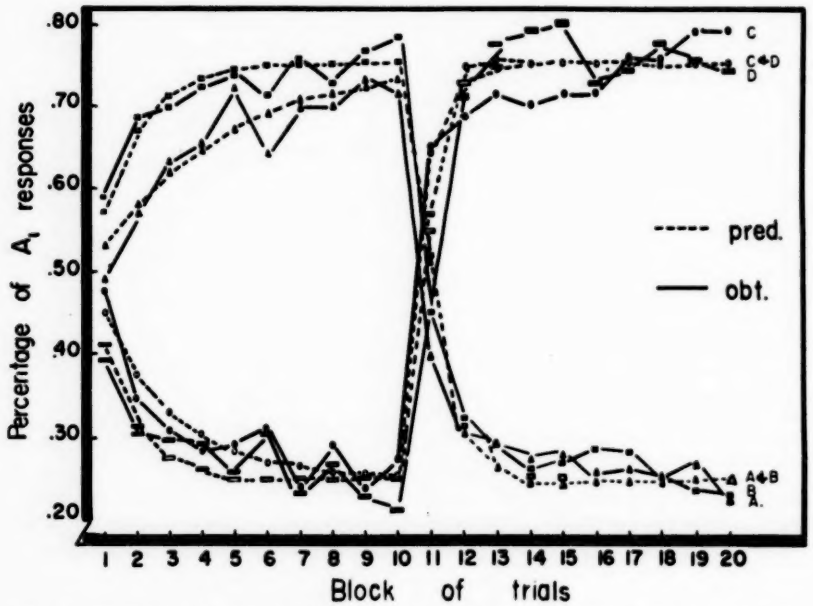


FIG. 1. Predicted and obtained marginal response probabilities for blocks of 24 trials (Exp. I).

Analyses of variance of mean marginal response probabilities for Groups A and B and for Groups C and D for the first and second 240 trials were computed. The only significant F ($P < .01$) was that for Groups A and B for the first 240 trials.

Blocks of the first 24 trials following a change in conditional probability specification without a change in marginal probability specification provide information concerning the dependence of marginal response probabilities on conditional stimulus probabilities. If marginal response probability were in fact directly related to conditional stimulus probability, the deviation of predicted minus obtained marginal probability values would become more positive when $\pi(1|1)$ increases and more negative when $\pi(1|1)$ decreases. The observed changes are in the direction required by independence in half of the eight cases and in the opposite direction in the other half. Although the average magnitude of change in the direction required by dependence is about twice as large as that in the direction required by independence, the result would seem to offer some empirical justification for the fact that Equation 1 does not depend on $\pi(1|1)$.

In general, the tabulated value of α decreased from the first block of 240 trials to the second block of 240 trials, regardless of whether comparisons are made for the same *S*s across variable conditions or for different *S*s across the same conditions. At least two explanations for this are possible. It may be a learning-how-to-learn phenomenon, α being an inverse measure of rate of learning. However, previous experimental results (**4**) suggest that α is directly related to the difference between initial and predicted asymptotic re-

sponse probabilities. Since in the present data this difference is twice as large in the first 240 trials as in the second, either explanation is admissible.

Note that α values for the first 240 trials for Ss who began learning with an r series followed by a d series were comparable as were values for Ss who had the series in the reverse order, regardless of marginal stimulus probabilities.

When marginal probability specification changed after 240 trials, dependency structure of the series changed in half of the cases and in half it did not. Tabulated α values for the second 240 trials were higher for groups in which this dependency change occurred than for groups in which it did not, regardless of the direction of change, marginal probability specifications, and α values for the first 240 trials.

Conditional data.—Predicted and obtained conditional response probabilities for blocks of 24 trials are shown in Fig. 2. Clearly, conditional response probabilities change from their initial values to more closely approximate conditional stimulus probabilities. However, it is not so clear that response probability eventually matches stimulus probability as it was in the marginal case. Asymptotic behavior of conditional response probabilities is discussed in more detail with reference to Exp. II where 240, rather than 120, trials were run without changing conditional stimulus probabilities.

In order to evaluate the model, pairs of predicted curves are shown in Fig. 2. The model requires that

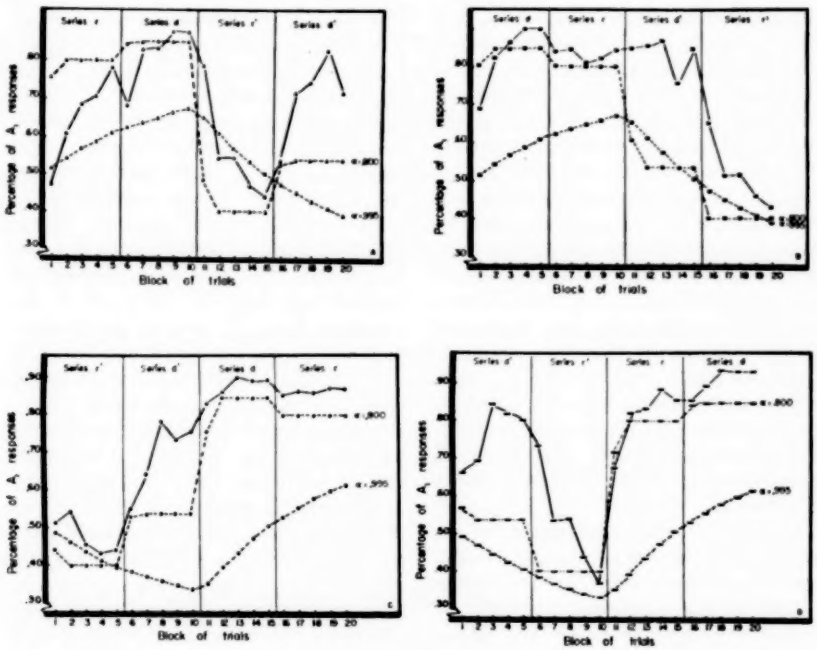


FIG. 2. Predicted and obtained conditional response probabilities for blocks of 24 trials (Exp. I).

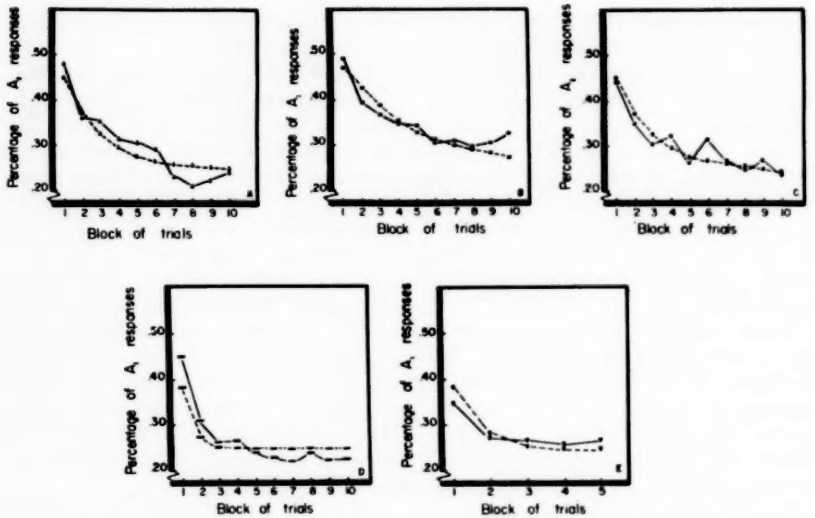


FIG. 3. Predicted and obtained marginal response probabilities for blocks of 24 (or 48) trials (Exp. II).

α values estimated for marginal data be appropriate also for the corresponding conditional data. This strict requirement is clearly not met; the question is, then, whether or not the model can predict the obtained data if a less severe requirement is imposed. In the present and previous experiments, α values for marginal data lie well within the range .800 to .995. (Actually values outside this range imply either instantaneous learning or virtually no learning.) These values were therefore taken as practical bounds on α and the corresponding predicted curves graphed. $V_{1,0}$ was taken as .50 for the first 120 trials for each group and obtained from Equation 1 for successive blocks of 120 trials.

In general, values obtained fail to fall within theoretical bounds. Exceptions deserve mention: for Series r and r′ marginal and conditional probability specifications are identical. Since conditional response probabilities tend to follow conditional stimulus specifications and theoretical prediction of asymptotic conditional response probabilities is bounded by marginal stimulus probability, it is reasonable to expect some correspondence between predicted and obtained values in these cases, at least asymptotically. With respect to the other case in which obtained values fell within theoretical bounds (Group A, Series d), it seems reasonable to say, especially on the basis of the results of Exp. II, that asymptotically this would not have been the case.

Marginal and conditional response probabilities are remarkably independent as can be seen from a comparison of Figs. 1 and 2. It seems clear, then, that any equation whose prediction of asymptotic conditional response behavior is bounded by marginal stimulus probability is inappropriate. Conditional response probabilities tend to follow conditional probability specifications of the series. If these specifications are taken as approximations of their asymptotes,

the values of α required by Equation 2a are 1.00 for Series r and r', .625 for Series d, and .484 for Series d'. These required α values, far outside the .800 to .995 range, are further evidence of the inappropriateness of Equation 2a.

Experiment II

Marginal data.—Predicted and obtained curves for blocks of 24 (or 48) trials are given for Groups A, B, C, D, and E in Fig. 3. For Groups F, G, H, I, and J a stationary response probability of .500 is predicted and approximately obtained. (Mean response probabilities ranged from .492 to .499, probabilities for blocks ranged from .471 to .521, and no systematic trends occurred.) The model again seems able to predict marginal learning curves quite accurately, even when these curves are straight lines.

An analysis of variance of mean marginal response probabilities was computed. Conditional stimulus probability failed to have a significant effect ($F = 2.24$; 2.41 required at .05 level), but the interaction of marginal and conditional stimulus probabilities was significant ($F = 3.10$; 3.41 required at .01 level). Conditional stimulus probability influenced marginal response probability for Groups A, B, C, D, and E, but not for Groups F, G, H, I, and J. The model requires that marginal response probability be independent of conditional stimulus probability; this requirement seems to be met when marginal response learning is not predicted (and does not occur), but not when such learning is predicted (and does, in fact, occur). Of course, generalizations must be made with caution since only two marginal probability specifications were used.

Estimated values of α used in plotting theoretical curves for Groups A, B, C, D, and E are given in Table 5. V_{L0} was taken as .50 in each case.[8] A V_{L0} of .50 for Groups F, G, H, I, and J gives an α of 1 in each case. Thus, as noted above, no change in initial response probabilities is predicted for the latter groups. Note that differences among the predicted learning curves for Groups A, B, C, D, and E are entirely attributable to differences in estimated α values.

TABLE 5

ESTIMATED VALUES OF α BASED ON MARGINAL DATA FOR EXP. II

Exp. Group	First 120 Trials	First 240 Trials
A	.982	.980
B	.985	.990
C	.976	.980
D	.963	.937
E	.962	.970

There is some evidence that α values vary as a function of difference between marginal and conditional probability specifications of the series. That is, as the difference in specifications increases, α in general decreases. This relationship apparently does not hold for Group E. The data of Exp. I are relevant if values are computed for only the first 120 trials for each group (Table 4). When this is done the relationship is seen to hold, although the test is limited since only two conditional specifications were used with each marginal specification. If α values are recomputed for Groups A, B, C, D, and E of Exp. II using only the first 120 trials (so that these estimates will be strictly comparable to those from Exp. I), Groups D and E fall nicely in line and, furthermore, differences in estimated α values are more nearly proportional to differ-

[8] Again .50 is the value assigned to V_{L0} a priori, and, since in no case did the empirical value differ significantly from .50, it was used for all groups.

ences between series specifications (Table 5). Mathematically, this is a result of the fact that obtained marginal response probabilities for Group D were consistently below the corresponding predicted values for the second 120 trials, but not for the first 120 trials. An estimation procedure for α which specifies the asymptote ($\pi(1)$) will thus give an appreciably lower estimate based on 240 trials than on 120 trials for such data.

Conditional data.—Conditional response probabilities are again remarkably independent of marginal response probabilities and tend to follow conditional stimulus probabilities, as may be clearly seen in Fig. 4. As in Exp. I, it is not clear that conditional response probabilities eventually match conditional stimulus probabilities. Actually conditional response asymptotes seem generally to exceed conditional probability specifications;[9] however, at least for some groups, these responses do not appear to be at asymptote after 240 trials and statements concerning asymptotes must therefore be tentative.

Predicted and obtained curves for blocks of 24 (or 48) trials are given in Fig. 5. $V_{1,0}$ is taken as .50 in each case. In general, empirical values fail to fall within theoretical bounds. Note that predictions for α values of .800 and .995 are shown for Groups F, G, H, I, and J, although the model clearly requires an α of 1 and thus a stationary conditional response probability (as well as a stationary marginal response probability) of .500 for these groups.

If conditional probability specification of the series is taken as an approximation of predicted asymptotic conditional response probability, the values of α required by Equation 2a are 1.00 for Groups B and H, .750 for Group C, .625 for Group D, .543 for Group E, .741 for Group I, and .556 for Group J. No admissible value of α can be computed for Groups A, F, and G. In general, it is im-

[9] A similar result has been noted by Anderson (1).

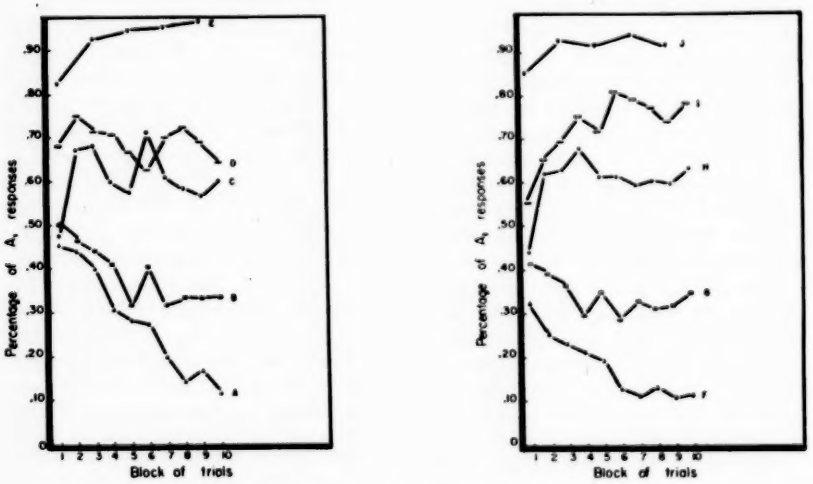


FIG. 4. Obtained conditional response probabilities for blocks of 24 (or 48) trials (Exp. II).

possible to compute α values satisfying Equation 2a for groups whose conditional probability specification is below marginal probability specification, if conditional specification is taken as asymptote. This, of course, follows from the fact that the asymptote of Equation 2a has $\pi(1)$ as lower

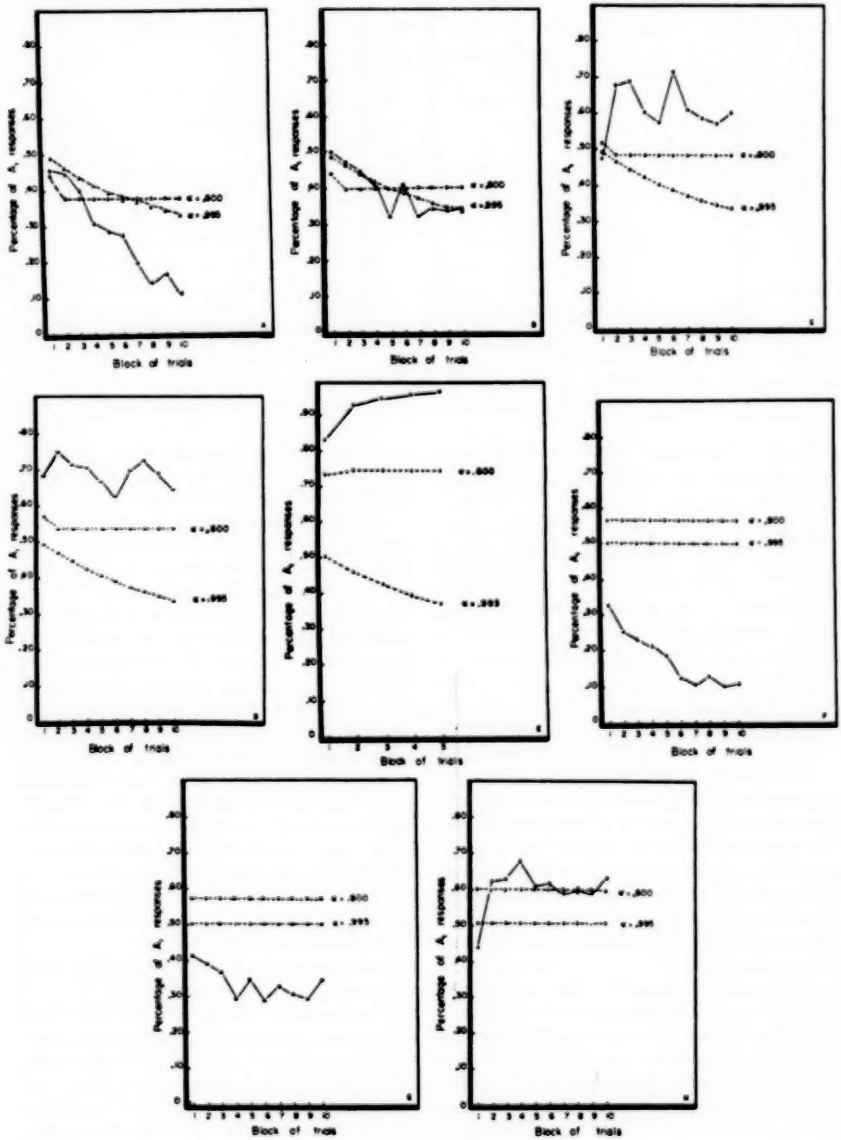


FIG. 5. Continued on next page.

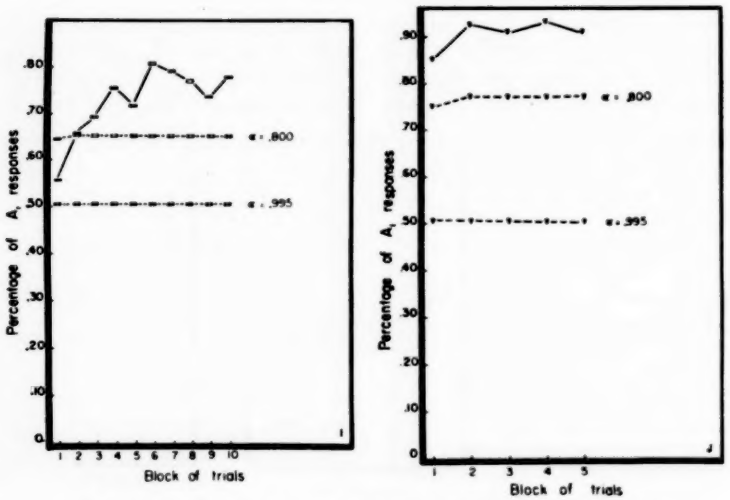


FIG. 5. Predicted and obtained conditional response probabilities for blocks of 24 (or 48) trials (Exp. II).

bound. Although precise empirical asymptotes of conditional response probabilities for Groups A, F, and G may be in doubt, they are clearly below marginal stimulus probabilities and are, furthermore, apparently independent of such specifications. This would seem to be conclusive evidence of the inappropriateness of Equation 2a.

DISCUSSION

The experiments of this study were designed to provide information concerning (*a*) the extent to which the Bush-Mosteller model can predict aspects of empirical data other than asymptotic marginal distributions of responses; and (*b*) the experimental conditions of which the learning parameter, α, can be shown to be a function.

With regard to the first question, previous investigators have found that, for the independent case, the model predicts accurately asymptotic mean marginal response probabilities and generates learning curves closely approximating the observed trends. The present experiments support these findings and show, for the first time, that they hold also in the Markov case.

However, the model requires that the asymptotic probability of Response A_1 following event E_1 depend on both the marginal probability of E_1 and on α. The results of the present study indicate that this probability depends mainly on the probability of E_1 following E_1. That *S*s' responses should thus follow the conditional structure of the experimental series as well as its marginal structure seems quite reasonable. Since *S* is trying to guess how the series will behave, it is unlikely that he should fail to discriminate between independent series and those with considerable sequential dependency among stimuli, as it is just this type of discrimination which will improve his prediction of the behavior of the series.

Since predicted conditional learning curves are clearly inappropriate, it is a moot question whether or not learning parameters computed for marginal data are appropriate also for conditional data. In effect, this means that no measure of rate of conditional learning is available until the equation for condi-

tional learning is specified. Some of the characteristics of the data which such an equation must predict will now be considered.

It is probably inappropriate to make statements about asymptotes of conditional responses in Exp. I. Any such statements require either that the theoretical asymptote be specified and approach to it considered, or that the data in question remain relatively stable for a sufficiently long period of time to be reasonably considered approximately at asymptote. The asymptote predicted by the model has been shown to be wrong, a satisfactory substitute has not been offered, and in no case does it appear that the data in question actually reach asymptote. Although this is actually an oversimplification of the data, suppose that conditional probability specification is taken as approximate conditional response asymptote as a reference point for discussion.

The *S*s apparently thought runs of E_1 and E_2 stimuli were longer than was actually the case or else they found such runs easy to learn,[10] since series with high $\pi(1|1)$ specifications were more rapidly learned regardless of the magnitudes of the required changes in response probabilities. Groups A and F of Exp. II, for which conditional probability specification was .10, are an exception. Conditional responses of some of the groups of Exp. II appear to be approximately at asymptote, but certainly this is not the case for all groups. Apparently, then, conditional learning is somewhat slower than marginal learning, since marginal response probabilities usually reach asymptote by the end of 120 trials (see Table 6). Furthermore, conditional learning rate seems to depend on marginal probability specification of the series (compare Groups A, B, C, D, and E with Groups F,G, H, I, and J). Generalizations must be made with caution since marginal learning was required for some of these groups, but not for others.

Note that not predicting E_1 is the same as predicting E_2. Furthermore, of course, Response A_1 follows either an E_1 or an E_2. Also, if *S* is to maintain a fixed marginal response probability, increased probability of Response A_1 following E_1 implies increased probability of Response A_2 following E_2. Thus *S*'s response on each trial contributes to both his marginal and conditional learning curves, at least indirectly. The *S*s show remarkable ability to alter their conditional response probabilities with or without altering their marginal response probabilities, depending upon series specifications. However, this does not necessarily imply that conditional learning rate is independent of marginal stimulus probability. In fact, marginal and conditional probabilities are always related by the condition

$$V_{1,n} = \pi(1)[V_{1,n}|E_1] + \pi(2)[V_{1,n}|E_2].$$

The data of the present study, then, do not provide parametric information concerning the apparent effect of marginal probability specification on conditional learning rate, and the information provided is difficult to evaluate since conditional learning rate has not been satisfactorily specified mathematically. With regard to the effect of conditional probability specification on marginal learning rate the situation is clearer. Dependency structure (independent versus Markov) of a series seems to influence rate of marginal learning independently of marginal stimulus specifications of the series in the following manner: (*a*) Markov series seem to be easier to learn than random series, which is reasonable since *S*s expect longer runs of stimuli than actually occur; (*b*) when marginal probability specifications are changed, learning is faster if dependency structure is not changed, regardless of what that structure is. Furthermore, learning rate seems to be directly related to the difference between marginal and conditional probability specifications of the series.

[10] The probability of an A_1 response on the second trial following an E_1 stimulus on the first trial was .542. Apparently, then, *S*s to some extent expected runs of stimuli prior to actual experience with the experimental series.

TABLE 6

NUMBER OF TRIALS REQUIRED BY EQUATION 1 TO REACH ASYMPTOTE

α	$\lvert V_{1,0} - \pi(1)\rvert = .25$	$\lvert V_{1,0} - \pi(1)\rvert = .50$
.800	25	28
.850	34	38
.900	52	59
.925	71	80
.950	108	121
.960	135	152
.970	181	204
.980	272	308
.990	550	619
.995	1100	1240

The major difficulty in evaluating these results is the lack of a significance test for differences in α values. In order to afford some idea of what differences in α values mean in terms of behavior, the number of trials required for Equation 1 to reach asymptote (to the third decimal place) for $|V_{1,0} - \pi(1)| = .25$ and $|V_{1,0} - \pi(1)| = .50$ is given for different values of α in Table 6. The matter is obviously somewhat arbitrary, but it seems reasonable to conclude that the trends noted do, in fact, represent more than random fluctuations.

Perhaps the most striking finding of the present study is the ability of *S*s to follow, independently, marginal and conditional probability specifications of series in making their responses. This fact must be taken into account in any revision of the model which is to predict data accurately.

SUMMARY

A total of 280 *S*s were run in two experiments on a modified Humphreys-type light-guessing apparatus. The experiments were designed to provide data suitable for detailed comparisons with deductions from the Bush-Mosteller model.

Empirical results are as follows:

1. Asymptotic marginal response probabilities approximately match marginal stimulus probabilities for both independent and Markov series.

2. For a fixed marginal stimulus probability, significant differences in marginal response probabilities are produced by variation in conditional stimulus probability, under certain conditions.

3. Marginal learning rate varies systematically as a function of (*a*) either stage of practice or difference between initial and predicted asymptotic marginal response probabilities (or both); (*b*) conditional probability specification of the series with respect to both dependency structure (independent versus Markov) and degree of dependency (in the Markov case).

4. Asymptotic conditional response probabilities (as approximated by behavior after 240 trials) appear to exceed conditional stimulus probabilities except when the latter are extremely low.

5. Conditional learning rate seems to vary as a function of marginal stimulus probability; however, the generality of this result is in question.

6. Marginal and conditional response probabilities change quite independently to follow the corresponding stimulus probabilities.

With regard to theory, the conclusions are as follows:

1. The model quite accurately predicts marginal learning curves and asymptotic response probabilities.

2. The model is unable to predict either conditional learning curves or asymptotic response probabilities. Learning rate is therefore not satisfactorily specified, but appears to be related to several experimental variables rather than to match marginal learning rate as required by the model. Conditional response probabilities below marginal stimulus probabilities cannot be predicted by the model; since conditional response probabilities tend to follow conditional stimulus probabilities, prediction of conditional response probabilities for groups where conditional stimulus probability is below marginal stimulus probability is therefore impossible for any admissible parameter value. A major theoretical revision with respect to conditional response probabilities seems required.

REFERENCES

1. ANDERSON, N. H. Effect of conditional probability and intertrial interval in a two-choice situation. Paper read at Midwestern Psychol. Ass., St. Louis, May 4, 1956.

2. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.

3. Engler, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. Unpublished doctor's dissertation, Northwestern Univer., 1957.
4. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
5. Goodnow, J. J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68**, 106–116.
6. Grant, D. A., Hake, H. W., & Hornseth, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.
7. Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.
8. Humphreys, L. G. Acquisition and extinction of verbal expectancies in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
9. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and a critique. *Psychol. Bull.*, 1951, **41**, 291–297.

(Received May 24, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

LONG-TERM RETENTION OF PERCEPTUAL-MOTOR SKILLS[1]

R. B. AMMONS, R. G. FARR, EDITH BLOCH, EVA NEUMANN, MUKUL DEY, RALPH MARION, AND C. H. AMMONS

University of Louisville[2]

It is commonly believed that perceptual-motor skills are well retained over no-practice intervals and are quickly relearned. Detailed information about the retention of such skills may be found in McGeoch and Irion (**5**). Of particular interest are studies reported by Tsai (**9**), Van Dusen and Schlosberg (**10**), and Jones and Bilodeau (**3**).

Several preliminary studies were carried out preparatory to the two major ones reported in the present paper.[3] There was no detectable loss of a procedural skill over a no-practice interval of 1 day (**2**), or of a compensatory pursuit skill over a no-practice interval of 1 mo. (**1**). On the other hand, half of the learned proficiency in performing another procedural skill was lost over a 2-day no-practice interval; and, after a 1-yr. no-practice interval, proficiency was little or no higher than on the first learning trial, with almost complete, even if transitory, loss of the skill (**7**). This indicates that not all perceptual-motor skills are well retained.

The purpose of the present studies was to measure the retention of two representative perceptual-motor skills following different amounts of initial training and following no-practice intervals up to approximately two years.

General Procedure

In order to increase the generality of the findings as much as possible, it was decided to use one task which called for a fairly complex type of compensatory pursuit skill (Airplane Control Test) and another which called for the sequential manipulation of a series of "controls." "Pursuit" and "following a sequence" were judged to be important components of a great many activities.

It was necessary to control variables that might be correlated with particular times during the 3-yr. training and retraining period. The principal problem areas were those related to apparatus, *E*s, and composition of groups. Preliminary work (**1, 2**) had shown that certain maintenance procedures were sufficient to keep the apparatus fairly constant in properties.

[1] This research was supported by the United States Air Force under Contract No. 18 (600)-356, monitored by the Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government. We wish to express our gratitude for the help given us by the following Air Force monitors of our project: Robert Gagne, Victor Montgomery, E. A. Bilodeau, and Jack Buel.

[2] The authors are now at the following locations: R. B. Ammons and C. H. Ammons, Montana State University; R. G. Farr, Denver, Colo.; E. Bloch, Louisville, Ky.; E. Neumann, Elmira College; M. Dey, Calcutta, India; and R. Marion, Bell Telephone, Louisville, Ky.

[3] In order to simplify the reading of this paper, we will suggest some rough definitions of frequently used terms. *Perceptual-motor* (or *motor*) *skill*—a temporally and spatially patterned movement of body musculature, the perception of whose accuracy and timing forms the basis of attempts toward maintenance or decrease of the discrepancy between the movement and the "ideal" pattern. *Procedural* (or *sequential*) *skill*—a perceptual-motor skill which calls for the making of a series of simple movements in a particular order. *Compensatory pursuit skill*—a perceptual-motor skill which calls for subject movements to match movements of a piece of equipment, where subject movements alter and thus partially determine those of the equipment.

The *Es*[4] were all carefully trained by one supervisor, and were constantly supervised by him in order to insure consistent treatment of *Ss*.

Particular attention was paid to the composition of the groups. Only male university students served as *Ss*. Inspection of early results indicated that there were no detectable differences between *Ss* associated primarily with ages (college classes), so the main control was the source.[5] An approximately equal proportion of each group came from the college of engineering, the physical education department, lower division general social science courses, psychology classes, and campus fraternities. The *Ss* were paid for work on the Airplane Control Test, but not for the procedural task, as the latter took much less time and was reported as more interesting and hence self-motivating.

Return of *Ss* for retraining was also an important problem. At the completion of training, a special effort was made to see that *S* released tensions and left with a good feeling, as this had been shown to be important in an earlier study in the laboratory (**4**). As many *Ss* as possible were persuaded to serve in both experiments.

Experiment I

Procedure

The equipment for studying the learning and retention of a procedural task has been described in detail elsewhere (**2**). It consisted of a large vertical panel on which 17 control units were mounted (e.g., a knife switch, a toggle switch, an automobile turn indicator, an automobile door handle, a rotary switch, a doorbell buzzer, a sliding door latch), so arranged that an *S* sitting in position in front of the board could easily reach each one. Only 15 of the items were actually used in the training and retraining. These items were arranged in three different sequences, and each *S* was assigned randomly to one of the sequences which was then used for him in both the training and retraining sessions. No significant differences in performance on the three sequences had been found in a preliminary study.

The correct sequence was indicated on a schematic chart hanging on the panel. When *S* manipulated an item incorrectly (manipulative error) or out of order (sequential error), a large red signal light, mounted in the top center of the panel, went out. The *S*'s task[6] was to correct himself and turn the light on before proceeding, working as rapidly as possible. Each time through the sequence was counted as a trial, and 1-min. rests were given between trials. Time and errors were recorded for each trial. Preliminary work had indicated that there would be little or no reminiscence with this duration of intertrial rest. It should also be noted that Reynolds and Bilodeau (**8**) found that condition of distribution during training had little effect on performance, during the retraining period, of several perceptual-motor skills.

In one of the preliminary studies, verbal reports were elicited from a number of *Ss* in order to obtain some idea of the nature of the learning. Reports were in agreement that the first few trials were used to learn to identify the items and their correct sequence, as well as how to manipulate them correctly. The next step was to speed up the run, still using the schematic chart as a guide. Finally, *S* learned to go quickly through the sequence from memory without looking at the chart. In a preliminary study (**2**) it had been observed that, when items were identified by number, some *Ss* used these numbers in learning the sequence. For the present study, a chart was used which showed the correct sequence in terms of simplified line drawings of the items themselves.

Groups of from 40 to 47 *Ss* (total $N = 538$) were trained under each of 12 conditions representing two degrees of learning (5 trials and 30 trials) combined with six durations of no-practice interval (1 min., 1 day, 1 mo., 6 mo., 1 yr., 2 yr.). The *Ss* were tested for retention and retrained for 10 trials after the no-practice interval. Theoretical and actual mean dura-

[4] In addition to the *Es* listed as authors of this report the following staff members also served as *Es* and as statisticians, and their help is gratefully acknowledged: Ruth Rozel, Criseyde Stone, Warren Robb, Charles Russell, Joel Bader, Lynn Curry, Ross Jordon, and Harvey Fishman.

[5] The assistance of the following individuals is gratefully acknowledged: Robert Warner, Dean Robert Ernst, John Heldman, A. W. Sroka, and I. B. Anderson. Without the cooperation of the whole faculty and student body of the University of Louisville these studies would not have been possible.

[6] Verbatim instructions for the procedural task and the compensatory pursuit task and a transcript of raw scores on both tasks may be obtained by ordering Document No. 5483 from the ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting in advance $95.20 for photocopies or $4.00 for microfilm. Make checks payable to Chief, Photoduplication Service, Library of Congress.

TABLE 1

PERFORMANCE OF MATCHED GROUPS ON PROCEDURAL TASK

No-Practice Interval		N	Means (Min.)			Mean Retraining Trials to Reach Final Training Level
Theoretical	Actual		First Training Trial	Last Training Trial	First Retraining Trial	
1 min.[a]	1 min.	35	1.728	.370	.443	1.61
24 hr.	23.8 hr.	36	1.714	.371	.495	2.19
1 mo.	30.8 da.	31	1.752	.393	.749	3.39
6 mo.	189.0 da.	28	1.710	.397	.796	3.39
1 yr.	355.5 da.	23	1.835	.395	1.267	3.88
2 yr.	704.4 da.	25	1.784	.406	1.294	4.68
1 min.[b]	1 min.	35	1.773	.189	.206	2.46
24 hr.	24.3 hr.	31	1.750	.214	.261	4.35
1 mo.	30.4 da.	35	1.779	.203	.544	8.11
6 mo.	180.8 da.	20	1.688	.206	.922	8.60
1 yr.	353.8 da.	25	1.750	.202	.831	9.67
2 yr.	695.4 da.	23	1.616	.202	.866	8.88

[a] Groups with 5 training trials.
[b] Groups with 30 training trials.

tions of no-practice intervals for matched groups are given in Table 1.

Matching

Groups were matched for time taken to complete the first trial of the learning session. The criteria used were that the F ratio for differences between means from a simple analysis of variance should be as small as possible, certainly not reaching significance at the 10% level; that the heterogeneity of within-groups variances should not reach significance at the 10% level by the Bartlett test; and that there should be equal N's in the groups. It was found impossible to meet the criteria and keep 30 *Ss* in each group. An alternate procedure was adopted, consisting of matching for means and variances and keeping as many *Ss* as possible. The general procedure followed was to eliminate more *Ss* from the groups which were extreme in mean or variance, to bring their means and variances into line with those of the other groups. Records were chosen for discard in such a way that the residual group mean and/or variance became more similar to those of the other groups. The N's after matching are shown in Table 1. They range from 20 to 36 a group, totaling 347. It was thus necessary to use statistical procedures taking the unequal N's into account.

Time scores on Trial 1 of the learning session, the matching trial, were organized into a 2×6 table, and the variance was analyzed. With the within-groups variance estimate as the error term, the F's for the other three estimates were in no case as large as .1. The Bartlett test gave $P > .50$, so the hypothesis of homogeneity of variance could not be rejected. To evaluate further the comparability of the groups, the median number of errors on Trial 1 for all groups was calculated and the number of *Ss* below and above this point was determined for each of the 12 groups separately, and χ^2 was computed. The obtained χ^2 of 4.67 with 11 *df* does not reach the 10% level of significance. Thus it can reasonably be concluded that the matching procedure was effective, and the matching satisfactory. No further use was made of error scores, since few errors were made by most *Ss* after the first one or two trials. For example, less than a quarter of the *Ss* made errors on the first retraining trial, after up to 2 yr. without practice. Mean time scores of the matched groups on Trial 1 are given in Table 1.

As a further rough check on the effectiveness of the matching, variance of the time scores of the six 5-trial training groups and the six 30-trial groups was analyzed for the last trial of their training periods (Trials 5 and 30, respectively). In both simple analyses of variance, $F < 1.00$ and the Bartlett χ^2 is such that $P > .50$. Thus, within the two training conditions, the hypothesis that the groups were representative of the same population could not be rejected. The distributions of time scores were highly skewed at these two points. The use of time scores on these last training trials as the basis for matching had been considered, but the great skewness and the small variability of scores weighed heavily against this procedure. Mean performance by the matched groups (Trials 5 and 30) is given in Table 1.

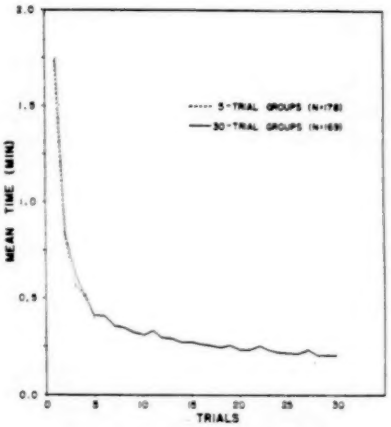


FIG. 1. Sequential task: performance during training period.

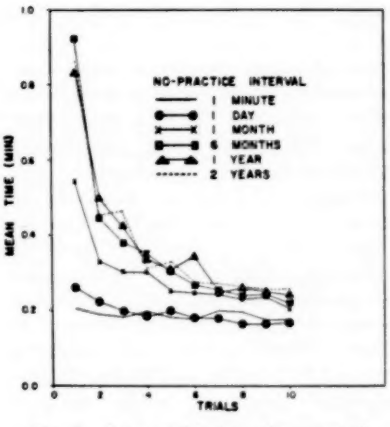


FIG. 3. Sequential task: performance of 30-training-trial groups during retraining.

Results

Figure 1 shows the mean time per trial during training for 5-trial and for 30-trial matched groups. It can be seen that a very rapid decrease in time per trial takes place during the first few trials, and that later in practice performance levels off.

Mean time per trial during retraining is shown in Figs. 2 and 3.

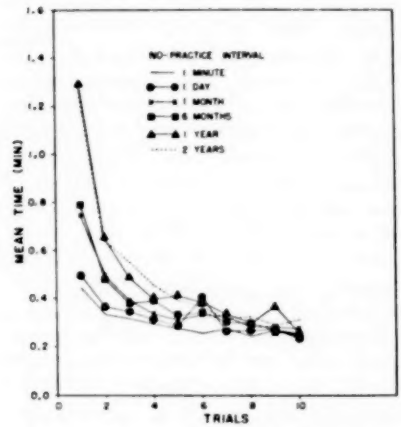


FIG. 2. Sequential task: performance of 5-training-trial groups during retraining.

In general the longer the no-practice interval the greater the loss in performance, and relearning is rapid. Mean performance on the first trial after the no-practice interval for the various groups is given in Table 1. With the scores arranged in a 2×6 table, the variance was analyzed. It was found that both amount of training and duration of no-practice interval had significant effects, but that their interaction was not significant (F's of 20.44, 26.38, and 2.03, respectively). The Bartlett test showed heterogeneity significant at far beyond the 1% level, the variance within groups being roughly correlated with the mean level of performance. For this reason, separate nonparametric tests were made for the groups with 5 and those with 30 training trials. The *Ss* in each of the 5-trial groups were classified as to whether they were above or below the Trial 6 median for the total 5-trial group. The χ^2 computed on the basis of this table was 78.52 with 5 *df*. The comparable χ^2 on Trial 31 for the 30-trial groups was 103.59. When less training is given

before the no-practice interval and there is a greater duration of that interval, absolute performance level is lower immediately after an interval of no practice. This is quite clear from the graphs, the analyses of variance, and the χ^2 analyses.

Number of trials to relearn to the level reached on the last training trial was computed for each *S*. Where that level had not been reached by the end of retraining (Trial 10), *S* was arbitrarily given a score of 11 trials. These scores were arranged into a 2 × 6 table, and their variance analyzed to give a rough estimate of the effects of amount of training and duration of no-practice interval. The *F*'s for both variables and their interaction were significant at far beyond the 1% level. The Bartlett test showed a heterogeneity significant at beyond the 1% level. Inspection of the group means given in Table 1 shows that the greater the amount of original training and the greater the duration of the no-practice interval, the more trials it took to regain the performance level reached before the no-practice interval was introduced. It took the 30-trial *S*s proportionally fewer trials to regain their earlier proficiency.

Estimates of percentage loss over the no-practice intervals were obtained. Each *S*'s time score on the last training trial was subtracted from his score on the first retraining trial, giving a loss score. His score on the last training trial was then subtracted from his score on the first training trial to give a gain-during-training score. Proportional loss over the no-practice interval was computed by dividing the loss score by the gain-during-training score. Mean percentage loss for the various groups is shown in Fig. 4. It can be seen that proportional loss, when

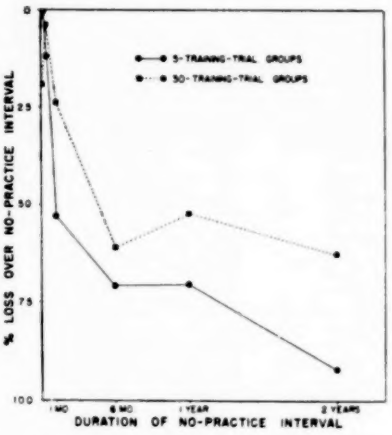


FIG. 4. Sequential task: performance loss over no-practice interval.

practice is interrupted, is greater the longer the no-practice interval and the less training given before that interval is introduced.

Inspection of Figs. 2 and 3 suggests that the effects of the longer no-practice intervals have not been completely overcome during the relearning, even after 10 trials. Analysis of variance (2 × 6 table) of the time scores at Trial 10 provides a rough check on this observation, even though score distributions are highly skewed and heterogeneous (Bartlett χ^2 of 129). The *F*'s for amount of training, duration of no-practice interval, and their interaction, using the within-groups variance estimate as the error term, are respectively 39.46, 8.24, and .92. Thus, the results of the analysis of variance are consistent with the observation.

Although no formal analysis was made, records of errors were inspected for clear trends. It was observed that there was an over-all reduction in number of errors paralleling the reduction in time per trial.

At the start of training, errors tended to be of the manipulative type, while later errors were almost all of the sequential type. After the no-practice interval, *Ss* tended to show increased sequential but not manipulative errors, at least on the first trial.

Experiment II

Procedure

The compensatory pursuit task (Airplane Control Test) used in this study has been described in detail elsewhere (**1, 6**). The *S* sits in front of a box housing a cam and pulley arrangement which moves a model plane mounted on a shaft above it and approximately at *S*'s eye level. The plane irregularly banks, climbs, and turns, going through four 1-min. cycles of movement before repeating its pattern. The *S* attempts to compensate for these movements, using hand (stick) and foot (pedal) controls. When the plane is level and pointed straight ahead, *S* is "on target," a small red light comes on, and the time during which this position is maintained is recorded. Score consists of time on target per 1-min. trial. Verbal reports of *Ss* indicated that the course of learning was approximately as follows: relations of plane and single control movements, then a coordination of the control movements into an integrated pattern, and finally an automatization of the patterned movements. Some *Ss* learned to sense the direction of movement of the plane from the feel of the controls, others did not. There were three test units, and *Ss* were trained as available, one to three at a time.

From 41 to 58 *Ss* (total $N = 465$) were trained under each of 10 conditions representing two degrees of training (1 hr. and 8 hr.) combined with five durations of no-practice interval (1 day, 1 mo., 6 mo., 1 yr., 2 yr.). The *Ss* were tested for retention and retrained for 2 hr. after the no-practice interval. Six additional *Ss* were trained for 40 hr. and retrained for 2 hr. after 1 yr. Theoretical and actual mean durations of no-practice intervals for the matched groups are given in Table 2.

The *Ss* were instructed[7] to keep the plane in level, straight-ahead flight as much of the time as possible. Rests of approximately 10 sec. were allowed between 1-min. trials, during which time scores were recorded. A rest of 5 min. was given after each 30 min. of actual practice. These arrangements had been favored over others by preliminary *Ss*, and led to little or no reminiscence or forgetting over 24-hr. periods (**1**).

During practice, *Ss* listened to music and engaged in casual conversation. During rests they could move around freely and do anything within reason except practice. Since 8-hr. training *Ss* ordinarily took more than one day to complete their training, a careful record was kept for the various groups with respect to the number of *Ss* taking various numbers of days to complete training. There was no evidence, however, that scores were appreciably affected by this variable. No 2-hr. period took more than 2 days, and no 8-hr. period more than 12 days to complete. The *Ss* in the 40-hr. training group were given approximately one month during which to complete their training.

Combination of Data and Matching of Groups

Since there was an unwieldy number of 1-min. scores, and since inspection indicated that no essential information would be lost, scores were combined by 15-min. periods. As this involved combining differing numbers of each of the four different 1-min. plane-movement cycles in different 15-min. periods, a check was made to see if there were any important differences in difficulty among the four cycles. Twenty *Ss* were found in the 8-hr. training groups whose performance curves during the last 60 min. of training were essentially flat. For each *S*, mean performance on each of the different 1-min. plane-movement cycles was computed. These means were then arranged into a Cycles by Subjects table (4×20). The variance was analyzed, giving *F*'s of 26.23 for Subjects and 2.08 for Cycles, using the variance estimate based on the residual as the error term. The *F* for Cycles does not reach the 10% level of significance with 3 and 57 *df*. It was concluded that cycles had a negligible effect, and could be safely disregarded in combining 1-min. scores into 15-min. scores.

The first step toward analysis was the matching of the groups. The criteria were that between-group variance should be less than within-group variance, that the within-group variances should not be significantly heterogeneous, and that the *N*'s should be equal, if possible. An attempt was made to match on the basis of time-on-target scores for the first 15 min. of training. Even with unequal *N*'s in groups, matching could only be accomplished by reducing some groups to fewer than 15 members. In spite of this wholesale discarding of records, simple analyses of variance of the time-on-target scores for the last 15 min. before the no-practice inter-

[7] See footnote 6.

TABLE 2

PERFORMANCE OF MATCHED GROUPS (N = 30) ON COMPENSATORY PURSUIT TASK (AIRPLANE CONTROL TEST)

No-Practice Interval		Means (% Time on Target)		Mean 15-min. Periods to Reach First Retraining Trial Level of Immed. Groups
Theoretical	Actual	First Retraining Trial	Last Retraining Trial	
24 hr.[a]	21.3 hr.	83.4	88.1	(control)
1 mo.	30.6 da.	83.5	88.1	1.03
6 mo.	189.2 da.	77.7	87.9	1.33
1 yr.	363.8 da.	75.2	87.8	1.70
2 yr.	706.1 da.	76.9	88.0	1.73
24 hr.[b]	21.6 hr.	93.3	90.3	(control)
1 mo.	31.4 da.	91.3	90.4	1.27
6 mo.	193.9 da.	90.7	90.0	1.57
1 yr.	373.8 da.	88.2	90.1	2.07
2 yr.	683.6 da.	85.9	90.1	3.20

[a] Four 15-min. practice periods during training.
[b] Thirty-two 15-min. practice periods during training.

val, for the 8-hr. and the 1-hr. groups taken separately, indicated that there were still significant differences in means and variances between the groups. When the groups were matched on the basis of scores during the last 15-min. period during training, 8-hr. and 1-hr. groups separately, differences between the groups on the first training trial were statistically significant, and variances were significantly heterogeneous.

Careful consideration of the problem led to the conclusion that scores during the training period before the no-practice interval could not be considered comparable for the various groups, and hence no matching on the basis of these scores would be satisfactory. A number of bits of information supported this conclusion or at least helped to rule out other factors. Systematic interviews with faculty members revealed agreement that morale among the men students had been lowest when the groups with lowest training scores had been run, due to such factors as the imminence of draft and war. Inspection of individual records showed that low scores during training were not necessarily followed by low scores during retraining. On the basis of comments by *Ss*, *all Es* agreed that performance on the Airplane Control Test seemed to be more affected by motivational differences among *Ss* than was performance on the procedural task, in part because practice lasted much longer. The *Ss* reported boredom much more frequently. It did not seem reasonable that the differences were in population, since its composition had been rather carefully controlled. In any case, approximately 50% of the *Ss* served in both experiments, yet the procedural task groups could be closely matched on the basis of training-period scores while the airplane control task groups could not be.

Since two different *Es* had run most of the *Ss*, one during the first year, and one during the second, it seemed possible that group differences might be due to *E* differences. Fortunately, it was possible to test this hypothesis, since each of the *Es* had run part of the *Ss* in each of two groups. Of the 1-hr.—6-mo. group, 19 had been run by E_1 and 24 by E_2. Of the 8-hr.—6-mo. group, E_1 had run 12 *Ss*, E_2 had run 27. Comparisons of group means on the first 15-min. retraining trial indicated that the *Ss* run by the two *Es* were closely comparable (*t*'s of .06 and .08, respectively, for the two groups).

It was finally decided to match the groups on the basis of their time-on-target scores for the last 15 min. of the retraining period, assuming that the differences between groups during training were in performance and not learning per se. The consistency of the results so obtained, and here reported, supports our reasoning. The five 1-hr. training groups and the five 8-hr. groups were matched separately for group means and variances. The means are given in Table 2, and are for an N of 30 in each group. The F's based on simple analysis of variance for the 1-hr. groups and for the 8-hr. groups were both less than 1.00, and the hypothesis of homogeneity of variance could not be rejected at the 50% level of confidence on the basis of the Bartlett tests (χ^2 equal to .15 and 1.40 with 4 *df*). It will be observed in Table 2 that the performance of the 8-hr.-

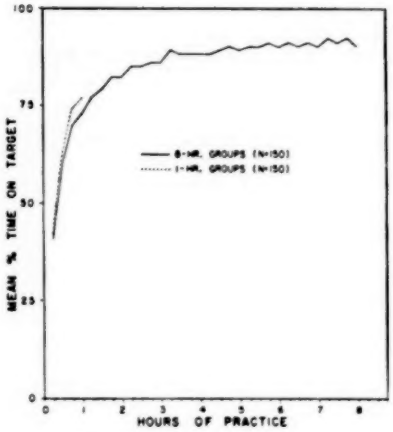


FIG. 5. Airplane control task: performance during training.

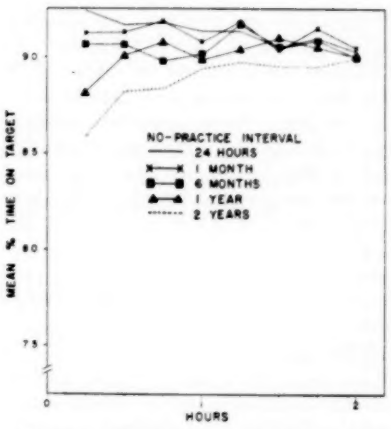


FIG. 7. Airplane control task: performance of 8-hr. groups during retraining.

training groups was consistently higher than that of the 1-hr. groups, which is reasonable in view of their having practiced nearly 10 hr. as compared with 3 hr. This difference is significant at well beyond the 1% level (F of 7.82 with 1 and 290 *df*, taken from over-all analysis of variance).

Results

Figure 5 gives the mean performance curves during training for all 1-hr. groups combined ($N = 150$) and for all 8-hr. groups combined ($N = 150$). It can be seen that the performance increases rapidly during the first hour or two, then gradually levels off, giving a typical negatively accelerated learning curve. The *S*s are on target about 40% of the possible time during the first 15 min. of practice, about 70% after 1 hr., and about 90% after 8 hr.

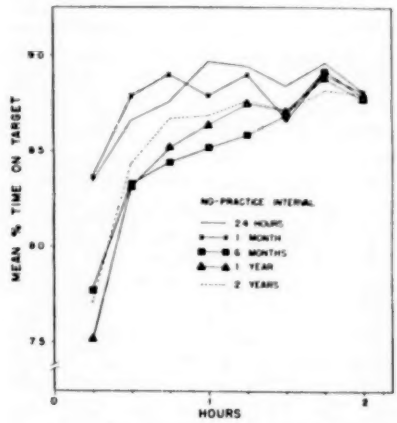


FIG. 6. Airplane control task: performance of 1-hr. groups during retraining.

Figures 6 and 7 give the mean performance curves for each no-practice interval group during the 2-hr. retraining period. Means for the first 15 min. of the retraining period are given in Table 2. Study of the curves and the means indicates that the longer the no-practice interval, the greater the loss (lower the performance at the start of retraining). The absolute loss over the 2-yr. period is about the same for the groups receiving 1-hr. and the groups receiving 8-hr. training. The variance (2 × 5 table) of percentage time-on-target scores for the first 15 min. of retraining was analyzed, and the results supported the above observations. The F's obtained, using

the within-groups variance estimate as error term, were 8.44 for no-practice intervals (*df* 4/290), 114.87 for duration of training before introduction of the no-practice interval (*df* 1/290), and 1.20 for the interaction (*df* 4/290). The within-group variances were significantly heterogeneous (Bartlett χ^2 of 80.00 with 9 *df*), but inspection showed that this heterogeneity was due primarily to the lower variability of scores in the 8-hr. groups as they approached a maximum possible score.

In order to give a clearer picture of what was happening to performance during the first part of the retraining, mean curves were plotted for each of the 10 groups by 1-min. periods during the first 15 min. These are shown in Figs. 8 and 9. It can be seen that the computed loss over the no-practice interval is greater than when only the combined 15-min. scores are used. This loss amounts to more than 20% time-on-target for one group when it is compared with its control group.

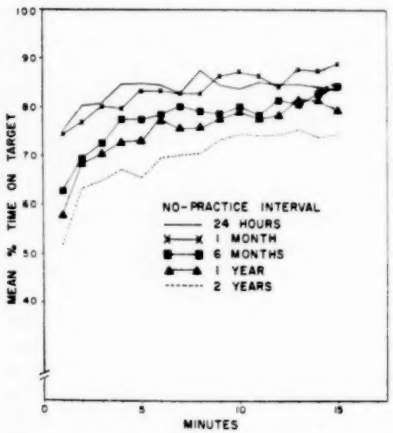


FIG. 8. Airplane control task: performance of 1-hr. groups during first 15 min. of retraining.

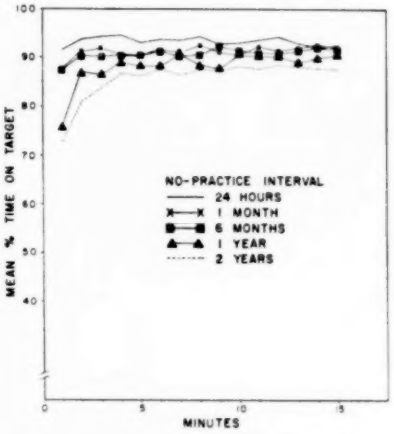


FIG. 9. Airplane control task: performance of 8-hr. groups during first 15 min. of retraining.

Proficiency is regained very rapidly, as much as 50% to 75% of the loss being recouped during the first 5 min. after resumption of practice. Loss tends to be greater the longer the interval without practice, as already indicated by the 15-min. scores.

Duration of retraining practice necessary to regain the level reached by the end of training (before the introduction of the no-practice interval) also provides an index of the degree of loss of proficiency over the no-practice interval. Due to inconsistency of performance during training, discussed earlier in some detail, this measure could not be computed. Instead, the mean scores during the first 15-min. retraining trial for the 1-hr. and 8-hr. groups with 24-hr. no-practice periods were used as a baseline, and the numbers of 15-min. periods necessary to reach these levels were computed for all *S*s, the 1-hr. mean being used for 1-hr. *S*s and the 8-hr. mean for 8-hr. *S*s. Those *S*s not reaching the control

levels were arbitrarily assigned a score of 9 (one more than the total number of 15-min. periods). This index of time to relearn has the advantage of taking rate of improvement into consideration, which is important with the airplane control task, since *Ss* are still improving rapidly after 1 hr. of practice but not after 8 hr. The means for the various groups are given in Table 2.

It can be seen that "relearning" takes longer, the greater the duration of the no-practice interval, and the greater the amount of training before introduction of that interval. However, the 8-hr. groups save only about 20 min. out of 1 hr. The absolute loss is greater but the relative loss is less in the case of the 8-hr. *Ss*. As a rough statistical check on the differences among groups, the individual scores (15-min. periods for *Ss* to reach initial retraining levels of the control groups) were arranged into a 2×5 table, and their variance was analyzed. The effects of retention intervals and amount of training were statistically significant, but their interaction was not (F's of 4.74, 3.77, 1.12, with df's of 4/290, 1/290, 4/290).

To obtain an estimate of the effect of a high degree of overlearning on retention, 6 *Ss* were given 40 hr. of training, and retrained after a no-practice interval of 1 yr. These *Ss* were carefully matched with 6 *Ss* from the 1-hr.—1-yr. group and 6 *Ss* from the 8-hr.—1-yr. group on the basis of percentage time on target during the first hour of training. The 1-hr., 8-hr., and 40-hr. means were 58.0, 57.9, and 57.9% for the 1-hr. matching period. After the 1-yr. no-practice interval, the respective means were 74.5, 91.4, and 91.0 for the first 15-min. retraining period. The t's were 5.40, 5.27, and .13 for the 1–8, 1–40, and 8–40 comparisons, with 10 df in each case. The first two t's are significant at beyond the 1% level. It is clear that as far as the time-on-target measure of proficiency was concerned, overlearning was not detectably helpful.

Summary

Twenty-two groups, ranging in size from 20 to 36 male college students, were trained to either a moderate or a high level of proficiency on a procedural task or on a compensatory pursuit task, then retrained after no-practice intervals up to 2 yr. in duration. Learning, retention, and relearning were measured. It was found that, on the bases of mean time taken per trial or percentage time on target per trial, absolute loss in level of proficiency was apparently not affected by amount of training and was greater the longer the no-practice interval. A greater *proportion* of proficiency was lost by groups receiving less training and groups receiving longer no-practice intervals. Retraining to the earlier level of proficiency took more trials the longer the no-practice interval and the greater the amount of training.

References

1. Farr, R. G., Dey, M. K., & Bloch, E. The Airplane Control Test: a compensatory pursuit task. *Percept. mot. Skills*, 1956, **6**, 77–80.
2. Farr, R. G., Russell, C. H., & Marion, R. Apparatus for studying serial motor learning. *Percept. mot. Skills*, 1956, **6**, 93–96.
3. Jones, E. I., & Bilodeau, E. A. *Retention and relearning of a complex perceptual-motor skill after ten months of no practice.* San Antonio, Tex.: Human Resources Research Center. Lackland Air Force Base, June, 1953. (*Res. Bull.* 53–17).
4. Lowe, W. F. Post-experimental discussion and attitude toward psychology. *Percept. Mot. Skills Res. Exch.*, 1952, **4**, 145–150.
5. McGeoch, J. A., & Irion, A. L. *The psychology of human learning.* (2nd ed.) New York: Longmans, Green, 1952.
6. Melton, A. W. (Ed.) *Apparatus tests.* AAF Aviation Psychology Program

Research Report No. 4, Restricted Supplement (Declassified). Washington, D. C.: U. S. Government Printing Office, 1947.

7. NEUMANN, E., & AMMONS, R. B. Acquisition and long-term retention of a simple serial perceptual-motor skill. *J. exp. Psychol.*, 1957, **53**, 159–161.
8. REYNOLDS, B., & BILODEAU, I. M. Acquisition and retention of three psychomotor tests as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.
9. TSAI, C. A comparative study of retention curves for motor habits. *Comp. psychol. Monogr.*, 1924, **2**, No. 11.
10. VAN DUSEN, F., & SCHLOSBERG, H. Further study of the retention of verbal and motor skills. *J. exp. Psychol.*, 1948, **38**, 526–534.

(Received October 23, 1956)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

MEANING, FREQUENCY, AND VISUAL DURATION THRESHOLD [1]

JANET A. TAYLOR

Northwestern University

One of the variables consistently shown to be related to the visual duration threshold of verbal material is the familiarity of the stimuli employed, familiarity typically being defined in terms of frequency of usage as determined by some type of word count (**3, 5, 9, 10**). Degree of familiarity or frequency of usage of words, however, has two highly correlated aspects, either or both of which may be variables contributing to its effects on threshold. First, words vary with respect to the sheer frequency of exposure to the letter pattern that an individual has experienced in the past. Second, such words may be assumed to vary in meaningfulness,[2] the more familiar, frequently used words possessing more meaning for the individual. Although this particular distinction does not appear to have been made in any discussion of the relationship between frequency of usage and threshold, an implicit assumption that sometimes seems to be made is that frequency per se is at least the principle factor involved (e.g., **8**). In general treatments of perception, on the other hand, meaningfulness seems to be assumed to be a variable influencing such response measures as duration threshold (e.g., **1, 7**).

A study by Solomon and Postman (**10**) indicates that frequency of exposure per se does affect threshold. These investigators had *Ss* read and pronounce unfamiliar Turkish words a varying number of times in a preliminary training period. They found that tachistoscopic recognition thresholds for these words were inversely related to frequency of prior exposure. The question that remains, then, is whether or not degree of meaningfulness is also a variable contributing to word threshold in addition to sheer frequency.

In the present study an attempt was made to equalize frequency of prior exposure of the experimental materials and to vary meaningfulness by experimental manipulation. Basically this was accomplished by showing two groups of *Ss* a list of nonsense syllables an equal number of times. One group studied only the syllables and the other saw each syllable accompanied by a picture of a common, familiar object which was designated as its "meaning." Assuming that the latter procedure is adequate to instill some meaning into the nonsense syllables, a comparison of the recognition thresholds of the two groups for the experimental materials should give some indication as to the role of this variable.

[1] Appreciation is expressed to Jane Courten for assistance in testing the *Ss* and to Jean Engler for statistical aid.

[2] Meaningfulness is being used here in a rather narrow, common-sense manner: an association between a verbal label and an object, event or concept with which *S* has had some experience. Thus, such usages of "meaning" as those referring to the relevance of the word to some contemporary motivational state, value system, etc., are being excluded. Presumably words may be described as having different degrees of meaningfulness to an individual due to such factors as the strength of the association between the word and its referent, the amount of experience with the latter, etc.

The validity of the above procedure rests, in part, on the assumption that the nonsense syllables have no appreciable meaning for the frequency group (those shown only the syllables). It is possible, however, that meaningful associations are elicited by such syllables per se and further that the strengths of these associations are equivalent to those produced experimentally for the meaning group. If this were the case, the meaning and frequency conditions would not be expected to result in any threshold differences, even if meaning were a variable.

In order to control for this possibility the basic experimental design was replicated with three different pairs of groups. Each pair was given a different set of nonsense syllables as stimuli, the sets varying in association value from low to moderate. The number and/or strengths of meaningful associations evoked by the nonsense syllables in a group can reasonably be assumed to be related to association value. Thus, if meaningfulness is a variable contributing to threshold, an inverse relationship between threshold and association value of the sets of syllables might well be found. Should this relationship occur, it would in itself suggest that meaningfulness plays a role in determining duration thresholds, independent of the results of the comparison between the frequency and meaning groups. Should no evidence be found that indicates that association value is related to threshold, it would seem difficult to ascribe any lack of difference between the frequency and meaning groups, were this to occur, to the obscuring effects of meaning already possessed by the syllables per se.

Method

Subjects.—Six groups were tested, each consisting of 20 *Ss* drawn from introductory psychology classes at Northwestern University. Membership in the groups was determined by a counterbalanced order of assignment as *Ss* appeared in the laboratory. Nineteen additional *Ss* were also tested but their records eliminated for reasons to be discussed later.

Materials.—Three sets of 10 nonsense syllables were employed, being drawn respectively from the 0%, 33% and 60% association lists prepared by Glaze (2). In each set the beginning consonants were never duplicated and, in general, formal intralist similarity was low. Because individual letters vary in the frequency with which they occur in our language and also differ in the ease with which they are discriminated as stimulus patterns, both factors contributing to threshold, an attempt was made to choose the syllables in such a manner that the letters involved occurred with equal frequency in each of the three sets. This was accomplished in large measure, equality being achieved in most instances and no discrepancy for a single letter between any pair of sets exceeding one. The syllables were also selected as to be easily pronounceable. Each of the three sets was further subdivided into two lists of five syllables each, one list being designated as the experimental materials and the second as control items used only during the threshold testing. Care was also taken to form these smaller groupings in such a manner that individual letters were equally represented in the two lists.

The syllables on the appropriate experimental list were presented to each *S* during training trials by means of 5 × 8-in. flash cards. For the frequency groups, each card had a single syllable typed on it in large block letters. For the meaning group, each card also contained a color picture of an object, the nonsense syllable appearing directly below it. The pictures, obtained from magazine illustrations and of approximately the same size, were all common, easily identifiable objects with one-syllable English names (e.g., chair). The same pictures were used with all three experimental lists. The three sets of syllables and the names of the objects paired with items on the experimental lists for each meaning group are reproduced in Table 1.

Apparatus.—Duration thresholds were obtained by means of a Gerbrands Mirror Tachistoscope. Both the exposure and pre-exposure field consisted of a white rectangle measuring

8 × 8 in. The fixation points, centered on the rectangle, were two small black dots placed just above and below the area where the stimulus appeared. The luminance of both fields was 1.0 ml., lighting being accomplished by means of fluorescent tubes. The stimulus items were typed in capital letters by an IBM electric typewriter (Executive model) which, unlike typical typewriters, automatically produced equal spacing between the letters.

Training trials.—All *Ss* were told that the study was part of a larger experiment investigating methods of learning foreign languages and that they would be studying "non-English words." They were further instructed that they would later be expected to know which "words" they had studied and, in the case of the meaning *Ss*, who saw each syllable accompanied by the picture of an object, the "meaning" of each word.

Each of the five flash cards from the appropriate list was exposed manually by *E* for approximately 2 sec. After all five cards had been exhibited, the cards were shuffled and the list presented again. This procedure was repeated until the set of cards had been shown 15 times. On each exposure, *S* was required to pronounce the nonsense syllable out loud. The treatment of the groups differed only in the cards utilized, the frequency groups seeing only the syllables and the meaning groups the syllables accompanied by the pictures.

Test trials.—After being given practice on the tachistoscope with an English word, *S* was informed that he would be shown non-English words, some of which he had just studied and the rest of the same type but unfamiliar. Eleven syllables were then exhibited, the first a practice or adaptation item whose threshold was not considered in the results and the remainder the appropriate experimental and control lists.

Each syllable was exposed first at .01 sec., the time being increased in .005-sec. steps until the syllable was correctly identified. The order of presentation of the syllables was random for each *S*. The *Ss* were encouraged to guess on each exposure.

An occasional *S* was extremely variable in his performance, often because of particular difficulty in recognizing a single syllable. Since the records of such *Ss* have a marked effect on means and variability, they were eliminated according to the following criterion: The range of threshold values within the control or the experimental list must be greater than .07 sec. The 19 *Ss* thus discarded came from all six groups, with approximately equal frequency. Fifteen of these *Ss* were eliminated on the basis of their performance on the control list, two on the experimental list, and two on both.

Memory tests.—Immediately following determination of thresholds, *Ss* were tested for their memory of the experimental materials. Each *S* was given, one at a time, 14 cards with single nonsense syllables typed on them. These consisted of the experimental and control lists from the appropriate set and 4 new syllables. For each item independently, *S* was asked to state whether or not it had been one of those studied during the training period. The *Ss* in the meaning group were given the five experimental syllables one at a time and, for each one, asked to name the object that had been associated with it.

Results

Duration thresholds.—The study formed a 2 × 3 × 2 factorial design: two conditions of prior presentation (frequency vs. meaning), three sets of materials (low, medium, and high

TABLE 1

Control and Experimental Lists and Names of the Objects Accompanying Experimental Items for the Three Sets of Material of Differing Association Value

Experimental Lists				Control Lists		
Low	Medium	High	Objects	Low	Medium	High
CEF	CEV	PAB	Cat	TOV	FOH	BEM
YOP	YOD	TAV	Spoon	DAX	DAJ	FOZ
WUB	BIW	JUK	Chair	MEF	MIP	DUX
GAH	GEP	WEC	Shoe	JID	WOX	CED
ZUK	TUZ	GIP	Lamp	VEC	KEC	HIJ

TABLE 2

MEAN THRESHOLDS (IN 1/100 SEC.) FOR THE SIX GROUPS

Condition	Control List						Experimental List					
	Low		Medium		High		Low		Medium		High	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Meaning	3.71	1.07	4.07	.59	4.05	1.20	3.45	1.10	3.43	1.01	3.24	.81
Frequency	3.78	1.33	4.01	1.08	3.64	1.30	3.26	1.03	3.45	1.01	2.93	.93

association value) and two lists of syllables (the previously studied experimental list and the control items). The mean threshold values for each of these cells are reported in Table 2 while the results of an analysis of variance for correlated measures are summarized in Table 3.

The control syllables, introduced to *S* for the first time during the tachistoscopic procedure, were included primarily to reduce the influence of pure guessing on the threshold measurement of the experimental items. The *F* comparing the thresholds of these two types of syllables is highly significant, thus confirming the well-known facilitating effect of prior knowledge of the stimuli.

Turning to the effect of the three degrees of association value, the *F* for this dimension was insignificant. The lack of effect is found in both the control and experimental lists as may be seen from the insignificant interaction between syllable list and set of materials and from inspection of the means in Table 2. Thus "meaning" in the sense of degree of association value of the nonsense materials cannot be said to have contributed to threshold.

Finally, the *F* for the meaning and frequency conditions was insignificant as were the interaction terms involving these conditions.

Memory tests.—Following the determination of thresholds, it will be recalled, all *S*s were tested for their memory of the materials involved during the training period. In the total group, all but three *S*s were able to identify without error which of the syllables they had studied and which they had not. Of the three *S*s whose identification was not perfect, two made one error out of a possible 14 and one made five errors. Fifty-five of the 60 meaning *S*s were also able to recall perfectly the names of the pictures associated with each stimulus. Four of the remaining *S*s made one error and one made two errors.

TABLE 3

ANALYSIS OF VARIANCE FOR CORRELATED MEASURES ON THE EXPERIMENTAL AND CONTROL LISTS FOR THE SIX GROUPS

Source	*df*	*MS*	*F*
Set (Assoc. Value)	2	1.60	.81
Condition (Freq.-Meaning)	1	1.28	.65
Set × Condition	2	.66	.34
Error (*b*)	114	1.97	
List (Exp.-Control)	1	20.59	89.91*
List × Set	2	.67	2.92
List × Condition	1	.01	.04
List × Condition × Set	2	.20	.85
Error (*w*)	114	.23	

* $P < .01$.

Discussion

While some may question the effectiveness of the experimental procedure used to induce meaning in the present study or protest that it was a pallid procedure when compared to real life, the lack of difference between the meaning and frequency conditions cannot reasonably be attributed either to the failure of the meaning groups to learn the English equivalents of the stimuli or to the obscuring influence of previously established "meaning" possessed by the nonsense syllables per se. Further, the very fact that the association value of the syllables had no demonstrable effect on threshold casts additional doubt on the role of meaning in contributing to threshold when other evidence is considered. In contrast to the results of the present study, association value *is* a variable contributing to performance in verbal learning (**4**) as is degree of meaningfulness of paralogs and actual words (**4, 6**). The different sets of materials, then, if not the experimental procedure used to induce additional meaning, should have been found to be related to threshold in the present study, if meaning were indeed a variable. Thus we may conclude from the present study that if meaning, as it is here defined, does have any effect at all on duration threshold, it is relatively weak in its effect and/or must be quite well established before its influence is evidenced. In contrast, sheer frequency of exposure to a stimulus pattern does play an important role in determining threshold; 15 exposures of the experimental syllables *did* lead to a highly significant difference between these and the control syllables even though the association value of the syllables and the addition of "meaning" to the stimuli by experimental manipulation had no effect. Previous research (**10**) has shown that even one previous exposure to a stimulus pattern has considerable effect on threshold, thus further indicating the marked influence of frequency in contrast to the weak or ineffectual meaning variable.

Summary

An attempt was made to determine whether meaningfulness is a variable contributing to the lower duration thresholds of familiar, frequently used words in addition to the factor of frequency of prior exposure to the stimulus pattern per se. The basic design consisted of giving one group of *Ss* a preliminary training period, during which *Ss* were shown, 15 times, a list of five nonsense syllables. A second group was also given 15 exposures of the list but saw each syllable accompanied by a picture of a common familiar object which they were asked to associate with it, i.e., to learn its meaning. This design was replicated with three different groups, each group being given different sets of nonsense syllables, varying in association value from low to moderate. Testing for tachistoscopic recognition of the experimental syllables and five additional control items of equal association value revealed that while all groups had a significantly lower threshold for the experimental list than for the control items, neither the degree of association value of the syllables nor the type of prior treatment (frequency vs. meaning) had any significant effect upon performance. A later test of recall showed very few errors in identifying the items studied during the training period and in remembering their "meaning." Thus the variable of meaningfulness, as defined by the procedures used here, was not shown to influence visual duration threshold.

References

1. Allport, F. H. *Theories of perception and the concept of structure.* New York: Wiley, 1955.
2. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35,** 255–269.
3. Howes, D. H., & Solomon, R. L. Visual duration threshold as a function of word probability. *J. exp. Psychol.*, 1951, **41,** 401–410.
4. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
5. McGinnies, E. M., Comer, P. B., & Lacey, O. L. Visual recognition thresholds as a function of word length and word frequency. *J. exp. Psychol.*, 1952, **44,** 65–69.

6. Noble, C. E. The role of stimulus meaning (*m*) in serial verbal learning. *J. exp. Psychol.*, 1952, **43,** 437–446.
7. Osgood, C. E. *Method and theory in experimental psychology.* New York: Oxford Press, 1953.
8. Postman, L. The experimental analysis of motivational factors in perception. In *Current theory and research in motivation.* Lincoln, Nebr.: Univer. Nebraska Press, 1953.
9. Solomon, R. L., & Howes, D. H. Word frequency, personal values, and visual duration thresholds. *Psychol. Rev.*, 1951, **58,** 256–270.
10. Solomon, R. L., & Postman, L. Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 1952, **43,** 195–201.

(Received March 11, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

ACQUISITION AND EXTINCTION OF THE CONDITIONED EYELID RESPONSE FOLLOWING PARTIAL AND CONTINUOUS REINFORCEMENT [1]

WILLIAM F. REYNOLDS[2]

State University of Iowa

In classical defense conditioning of the human eyelid response, when the UCS (air puff) follows the CS only on some fraction of the total trials, the procedure is referred to as partial or intermittent reinforcement. This technique is to be contrasted with 100% or continuous reinforcement in which the UCS is paired with the CS on every trial. The majority of investigations employing these procedures have been primarily concerned with the effects of type of reinforcement schedule upon resistance to extinction of the eyelid response.

Although some early studies (**2, 5**) reported no difference in level of conditioning during acquisition between partial and continuous reinforcement groups, subsequent experiments (**1, 3, 4**) have been consistent in finding higher levels of response under 100% reinforcement conditions. Resistance to extinction has been found to be significantly greater following partial reinforcement schedules than continuous reinforcement schedules (**1, 3, 4, 5**).

It should be noted that a standard extinction procedure has been employed in the experiments cited above on eyelid conditioning. The critical aspect of the extinction series is the deletion of the UCS from the usual CS-UCS sequence so that the CS is presented alone on each extinction trial. Under these conditions, extinction of the conditioned eyelid response is typically very rapid.

That level of response during extinction may be maintained for a considerably longer period has recently been demonstrated by McAllister (**9**). This investigator, using an 80% reinforcement schedule, conditioned *Ss*' eyelid responses to a tone at a CS-UCS interval of 450 msec. These responses were then "extinguished" by presenting the UCS 2500 msec. after the onset of the CS for 100 extinction trials, as McAllister had previously shown (**8**) that little or no conditioning would take place at this interval. The relatively slow decline of the extinction curve was striking. McAllister pointed out that the retention of the UCS in the stimulus sequence, but at an interval which did not produce conditioning, maintained the drive level (*D*) of the *Ss* which would have been substantially reduced during the usual extinction procedure.

The inclusion of the UCS in the stimulus sequence during extinction probably serves two important functions. First, it maintains the *D* of *S*, since in the classical defense situation the source of *D* is assumed, in part, to be a function of the general level of emotionality which occurs as a result of previous administrations of the noxious UCS. Secondly, the presence of the UCS reduces the possibility that a shift in the "set

[1] This article is based on a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The author is indebted to Kenneth W. Spence who directed the investigation.

[2] Now at Western Michigan University, Kalamazoo, Michigan.

to receive the UCS," which presumably develops during acquisition, will occur during extinction.

The present study made use of a modified definition of partial reinforcement in which the UCS was presented on all conditioning trials. The 60% or partial reinforcement condition employed a CS-UCS interval of 500 msec. on 60 "reinforced" trials, and an interval of 2500 msec. on 40 "nonreinforced" trials. The 100% reinforcement condition employed the 500 msec. interstimulus interval on each of the 100 conditioning trials.

Following the acquisition trials, *Ss* from each reinforcement condition were assigned to one of two extinction conditions. One extinction procedure (I) provided for the administration of the UCS on each of 100 extinction trials at a CS-UCS interval of 2500 msec. The second extinction procedure (II) provided for the deletion of the UCS from the stimulus sequence on all extinction trials.

Method

Subjects.—A total of 80 *Ss* were selected from the introductory courses in psychology at the State University of Iowa. Those *Ss* who scored below the 20th percentile on the Taylor Anxiety Scale (**14**) were not used in the experiment since it was desired to obtain as high a level of conditioning as possible. Two *Ss* who gave CR's on pretest trials, and 5 *Ss* whose conditioning records were ruined by mechanical difficulties were eliminated. One other *S* was also eliminated because he met the criteria employed in previous studies from the Iowa laboratory defining a voluntary responder (**13**). The remaining 72 *Ss*—48 males and 24 females—were assigned at random to the partial (P) or continuous (C) reinforcement acquisition conditions with the stipulation that the proportion of males to females was 2:1.

Apparatus and method of recording.—The *S* was seated in a dental chair in a semi-soundproof, darkened room. The *E* was in an adjoining room where the stimulus controls and recording equipment were located. The potentiometer-polygraph system used to record eyelid movements and for presenting the UCS was similar to that used in a previous study by Spence (**11**).

The room in which *S* was located was illuminated by an overhead 40-w. bulb to a level of .04 apparent ft.-candles. The CS consisted in the illumination of a 7-cm. circular disc to a level of .11 apparent ft.-candles. The UCS was an airpuff of 2.0 lb./sq. in. pressure delivered to the right cornea. The onsets and durations of the CS and UCS were controlled by means of Hunter-Brown decade-type electronic interval timers (**6**).

Conditioning procedure.—Following the reading of instructions, designed to facilitate conditioning, all *Ss* were given three preliminary trials to the CS alone. One preliminary presentation of the UCS was then given and *S*'s eyelid reactions recorded for 30 sec.

The verbal signal "ready" was presented 1, 2, or 3 sec. before the predetermined time for the beginning of a trial. The intertrial interval was randomly varied between 15, 20, and 25 sec. with a mean intertrial interval of 20 sec. To prevent *Ss* from becoming fatigued, a 10-min. rest period was given following Acquisition Trial 90. During this period, *S* was removed from the apparatus and allowed to walk about in an adjoining hall.

For those *Ss* run under the partial-reinforcement condition, the order of presentation of reinforced and nonreinforced trials was randomly varied, subject to the restriction that no more than three reinforced nor three nonreinforced trials occurred in succession.

Results

Acquisition of the CR.—Figure 1 presents curves of acquisition of the CR in terms of percentage of responses occurring in successive 10-trial blocks. Inspection of the curves reveals that the continuous reinforcement schedule produced a consistently higher level of response than did the partial-reinforcement schedule. This result is in agreement with those of the majority of studies comparing partial- and continuous-reinforcement conditions when the UCS is omitted during the nonreinforced conditioning trials. The rise in level of response shown in both curves following the rest period after Trial 90 is presumably a result of the dissipation of fatigue

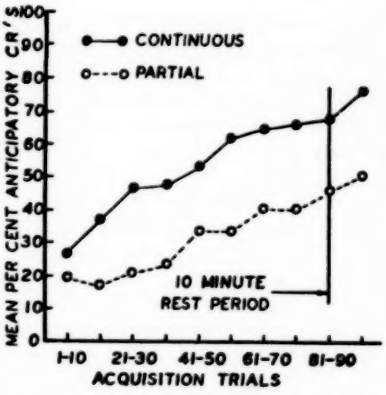


FIG. 1. Percentage of CR's on conditioning trials in 10-trial blocks.

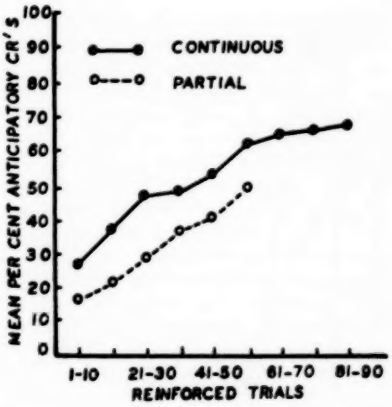


FIG. 2. Percentage of CR's on conditioning trials in blocks of 10 reinforced trials.

effects which had accumulated during the preceding conditioning trials.

The mean numbers of CR's exhibited during the 100 conditioning trials are shown in Table 1. An analysis of variance performed on these data resulted in an F of 15.22 ($P < .001$) for the main effect of Reinforcement Schedule. The F for Sex was 5.93, which was significant at the .05 level. This result is in line with those of previous experiments from the Iowa laboratory which have shown in general that female *Ss* tend to attain a higher level of response than male *Ss*. No evidence was obtained for an interaction between reinforcement condition and sex.

Figure 2 presents performance curves for the two reinforcement groups in which mean percentage of CR's is plotted as a function of successive blocks of 10 *reinforced* trials up to the 60th reinforcement. For *Ss* run under the partial-reinforcement condition these percentages were computed in the following way: The numerator of the ratio consisted of all CR's which occurred during a block of trials which included 10 reinforced trials; the denominator was the total number of trials in the block.

TABLE 1

NUMBER OF CR's OCCURRING IN 100 CONDITIONING TRIALS

Sex	N	Continuous Reinforcement		Partial Reinforcement	
		Mean	*SD*	Mean	*SD*
Male	24	49.63	5.49	27.38	4.46
Female	12	64.00	4.93	42.25	7.98

An inspection of Fig. 2 reveals that the 100% reinforcement *Ss* responded at a consistently higher level than the 60% reinforcement *Ss* when the number of reinforced trials was equated for both groups. The results of the analysis of variance performed on these data were essentially the same as the results of the 100-trial analysis. The F for Reinforcement Schedule was 4.47 ($P < .05$), and the F for Sex was 6.16 ($P < .05$).

Extinction of the CR.—Performance curves showing the percentages of anticipatory CR's as a function of successive blocks of 10 extinction trials are shown in Fig. 3. The block of trials designated 0 marks the final

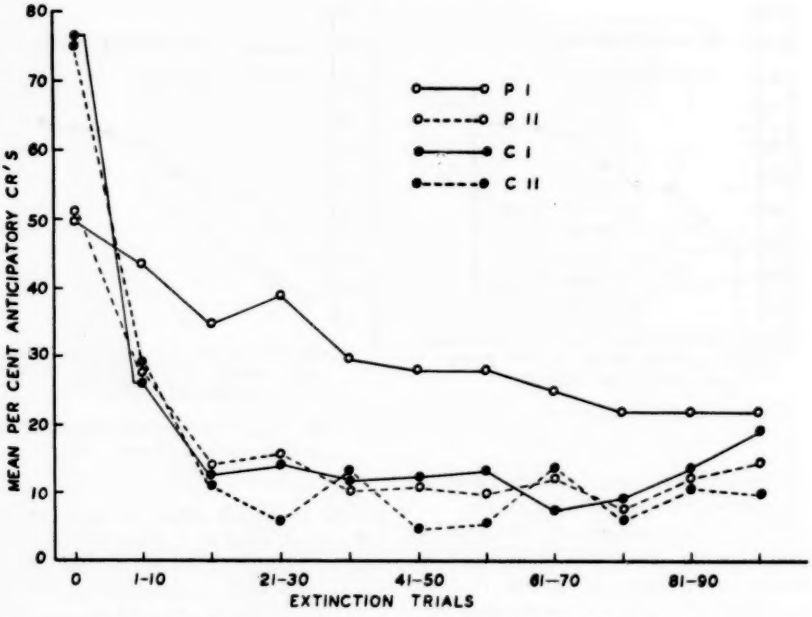


FIG. 3. Percentage of CR's on extinction trials in 10-trial blocks.

level of response reached during the conditioning trials (Trials 91–100). The parameters in Fig. 3 are the reinforcement conditions during conditioning, continuous (C) and partial (P), and extinction with the UCS present at the 2500 msec. interval (I) and the absence of the UCS during extinction (II). An examination of the figure shows that the extinction curve for the P-I group was consistently differentiated from the curves for the P-II, C-I, and C-II groups. Whereas the curve for Group P-I decreased gradually from the final acquisition level, the curves for the other groups dropped precipitously and showed little variation after the first block of 10 extinction trials.

Further inspection of Fig. 3 reveals that the presence of the UCS during the extinction trials had a differential

TABLE 2

ANALYSIS OF VARIANCE OF THE NUMBER OF ANTICIPATORY CR'S IN 100 EXTINCTION TRIALS

Source	*df*	*MS*	*F*
Reinforcement schedule (R)	1	157.25	5.69*
Extinction condition (E)	1	145.80	5.48*
Sex (S)	1	8.10	
R × E	1	73.47	2.76
R × S	1	32.40	1.22
E × S	1	24.02	
R × E × S	1	.34	
Error (*b*)	64	26.59	
Extinction trials (T)	9	21.80	13.71**
R × T	9	2.65	1.67
E × T	9	2.17	1.36
S × T	9	1.58	
R × E × T	9	1.90	1.19
E × S × T	9	.17	
R × S × T	9	1.51	
R × E × S × T	9	1.26	
Error (*w*)	576	1.59	

* $P = .05$.
** $P = .01$.

effect on the P groups but not on the C groups. The curves for the P-I and P-II groups show a marked and relatively consistent difference in level of response throughout the extinction trials. In contrast, the curves for the C-I and C-II groups follow nearly the same pattern during the extinction period with considerable crossing of the curves.

The results of an analysis of variance (7) carried out on these extinction data are summarized in Table 2. It may be seen that both the main effect of Reinforcement Schedule and the main effect of Extinction Condition were significant at the .05 level. The highly significant main effect of extinction trials ($P < .001$) refers to the fact that level of response declines during extinction. Neither the main effect of Sex nor any of the interactions involving Sex reached significance. The failure of any of the within-Ss interactions to attain significance means that no reliable differences were found between the slopes of any of the group curves.

The extinction data were also graphed when allowance was made for the fact that the extinction subgroups had attained different levels of response during the acquisition trials. The data for each subgroup were expressed, in 10-trial blocks, as percentages of the final performance level during conditioning. These "relative" curves of extinction did not differ markedly from the curves presented in Fig. 3 and for that reason are not presented here.

Discussion

The results of the present experiment are in agreement with those of the majority of previous studies involving aversive conditioning in showing that a higher level of response in acquisition is reached under conditions of continuous reinforcement than under partial reinforcement conditions. In the earlier studies, nonreinforcement involved the standard procedure of presenting the CS without the UCS. In the present study, nonreinforcement involved the presentation of the CS and the UCS on nonreinforced trials but at an interval (2500 msec.) that had been shown not to produce conditioning.

When the conditioning performances for the two reinforcement conditions were compared in terms of number of reinforced trials it was found that the curve for 100% reinforcement remained significantly above that for 60% reinforcement. This result would seem to suggest the presence of some inhibitory factor in the partial-reinforcement group which presumably developed as a function of the nonreinforcements.

The greater resistance to extinction of responses acquired under partial reinforcement than those acquired under continuous reinforcement has recently been given a stimulus generalization interpretation by Sheffield (10). According to this formulation, the stimulus aftereffects of a reinforced trial—in the present case the irritation resulting from the administration of the air puff—become part of the CS pattern acting at the start of the next trial. In like manner, the perseverative stimulus traces of nonreinforcement—the relaxation accompanying the omission of the UCS—also become part of the CS complex operating during the following trial. These perseverating cues of nonreinforcement are precisely those which follow the typical extinction trial. In partial reinforcement, then, the response becomes conditioned to a combination of reinforcement and extinction cues. From the principle of stimulus generalization, the partial-reinforcement CS will be more similar to the extinction CS than will the continuous-reinforcement CS and hence there will be less generalization decrement and more resistance to extinction following partial reinforcement.

The present experiment, by presenting the UCS on all conditioning trials, made

the stimulus aftereffects the same following reinforced and nonreinforced trials. The change from such uniform acquisition conditions to extinction conditions, with or without the UCS, should be the same for both reinforcement groups. Thus, with the stimulus traces of reinforcement and nonreinforcement made uniform and the change from conditioning to extinction stimuli made comparable, the Sheffield hypothesis would predict no difference between the continuous and partial groups in resistance to extinction under either extinction procedure.

This prediction was substantiated as far as the groups that were extinguished without the UCS were concerned. There was no difference in rate of extinction for the partial- and continuous-reinforcement groups under this condition; the curves for both groups showed a rapid decline to approximately a 10% level of response.

In the case of the groups extinguished with the UCS present at the 2500-msec. interval, however, the more general finding of greater resistance to extinction on the part of the partial-reinforcement group was obtained. This finding is not in accord with the Sheffield hypothesis. It should be noted in this connection that the continuous reinforcement *Ss* under this extinction condition did not exhibit the high resistance to extinction that might have been expected on the basis of the previous findings of McAllister (**9**) and Spence and Farber (**12**). Thus it will be seen that the curve of extinction for this group did not differ from that of the continuous group which was extinguished with no UCS. If the results for this group had been in accord with previous studies, no difference between the partial and continuous groups would have been obtained under this condition and both sets of data would have been in agreement with the Sheffield hypothesis. Just why the continuous-reinforcement group extinguished with the UCS present did not show greater resistance to extinction in the present study is not clear.

Summary

An experiment was conducted to provide data on the level of conditioning and resistance to extinction of the eyelid response following partial- and continuous-reinforcement schedules, when partial reinforcement was defined as including the UCS on all conditioning trials. The CS was the illumination of a milk-glass disc, and the UCS was an air puff of 2.0 lb./sq. in. pressure. Continuous-reinforcement *Ss* were given 100 conditioning trials with a CS-UCS interval of 500 msec.; partial-reinforcement *Ss* received 60 reinforced trials with an interstimulus interval of 500 msec., and 40 "nonreinforced" trials with a CS-UCS interval of 2500 msec. One-half of the *Ss* from each reinforcement condition were given 100 extinction trials with the UCS absent, while one-half of the *Ss* received 100 extinction trials with the UCS present at the CS-UCS interval of 2500 msec.

The continuous-reinforcement schedule produced a significantly higher level of response during conditioning trials than the partial-reinforcement schedule. The same result was obtained when the two reinforcement conditions were compared in terms of number of reinforced trials. Substantiating previous studies, female *Ss* reached a significantly higher level of response during conditioning than male *Ss*. Partial-reinforcement *Ss* extinguished with the UCS present, maintained a consistently higher level of response than those extinguished without the UCS. There was no difference during extinction between the continuous-reinforcement groups.

References

1. Grant, D. A., & Hake, H. W. Dark adaptation and the Humphreys random reinforcement phenomenon in human eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 417–423.
2. Grant, D. A., Riopelle, A. J., & Hake, H. W. Resistance to extinction and the pattern of reinforcement. I. Alternation of reinforcement and the conditioned eyelid response. *J. exp. Psychol.*, 1950, **40**, 53–60.
3. Grant, D. A., & Schipper, L. M. The acquisition and extinction of conditioned eyelid responses as a function of the percentage of fixed-ratio random reinforcement. *J. exp. Psychol.*, 1952, **43**, 313–320.
4. Grant, D. A., Schipper, L. M., & Ross, B. M. Effects of inter-trial interval

during acquisition on extinction of the conditioned eyelid response following partial reinforcement. *J. exp. Psychol.*, 1952, **44,** 203–210.

5. HUMPHREYS, L. G. The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25,** 141–158.
6. HUNTER, T. A., & BROWN, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62,** 570–575.
7. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.
8. MCALLISTER, W. R. Eyelid conditioning as a function of the CS-UCS interval. *J. exp. Psychol.*, 1953, **45,** 417–422.
9. MCALLISTER, W. R. The effect on eyelid conditioning of shifting the CS-UCS interval. *J. exp. Psychol.*, 1953, **45,** 423–428.
10. SHEFFIELD, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39,** 511–526.
11. SPENCE, K. W. Learning and performance in eyelid conditioning as a function of intensity of the UCS. *J. exp. Psychol.*, 1953, **45,** 57–63.
12. SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, **45,** 116–119.
13. SPENCE, K. W., & TAYLOR, J. A. Anxiety and strength of the UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42,** 183–188.
14. TAYLOR, J. A. The relationship of anxiety to the conditioned eyelid response. *J. exp. Psychol.*, 1951, **41,** 81–92.

(Received March 28, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

AN EXTENSION OF HULLIAN THEORY TO RESPONSE DECREMENTS RESULTING FROM DRUGS

R. B. PAYNE

Arctic Aeromedical Laboratory

One can hypothesize in molar fashion that the effects of depressant drugs upon learned behavior result in part from increased excitation thresholds in the neural structures which mediate response processes. Such threshold changes can be shown directly even if their intimate connections with response processes, at present, cannot. Within the framework of Hullian theory (**2**), one can suppose further that these induced increments summate physiologically with simple "inertial" thresholds so that, under given values of habit strength and drive, reaction potential becomes insufficiently suprathreshold to insure a high probability of response evocation. Thus, for a given value of reaction potential, response probability rises or falls as an inverse function of dosage.

From the foregoing theoretical model it follows that manipulations resulting in higher reaction potential will, under constant dosage, attenuate the degrading effect of a drug. Under the propositions (*a*) that reaction potential is a multiplicative function of habit strength and drive, and (*b*) that habit strength is a positive function of the number of reinforced trials, one way of increasing reaction potential would be to extend learning trials beyond those required to reach some initial criterion. This suggests the testable prediction that more highly learned responses will undergo less impairment from drugs than less highly learned responses. Moreover, the mitigative effect of "overlearning" should follow a law of diminishing returns, since successive units of training will yield progressively smaller increments of habit strength.

In order to test this proposition, Payne (**4**) trained groups of *Ss* on a serial order verbal task to criteria of 0%, 25%, and 50% overlearning, respectively, as defined by Luh (**3**), then administered clinical doses of antihistaminic and anticholinergic drugs along with appropriate placebo controls. Relearning trials conducted 75 min. posttreatment showed retention efficiency under the drug conditions as a negatively accelerated increasing function of the degree of learning, and showed all the statistically significant impairments as having occurred under the minimal condition. Thus, the prediction was confirmed both as to content and functional form. Drive was manipulated in the same experiment by the delivery of progress feedback to half the *Ss* in each subclass at the end of each relearning trial. Consistent with expectation, all the statistically significant drug impairments occurred under the no-feedback condition.

The rational model should work just as well, of course, for the prediction of outcomes resulting from operations designed to *lower* reaction potential from some previously established level. For example, consider the circumstance in which *Ss* during relearning are deprived of certain cues present during original learning. The original cues can be regarded as elements of a stimulus compound bonded to the same response through the resultant of their afferent interaction. Upon the withdrawal of certain cues during relearning, the residual cues, operating under the principle of the stimulus generalization gradient, will possess a joint power of reaction evocation smaller

than that commanded by the original compound. From this it follows that more trials will be required to relearn to criterion under the reduced cue condition than when all original cues are present. From the first experiment it follows, further, that if this event is made to occur in the presence of elevated thresholds resulting from drugs, even more trials will be required to relearn than under reduced cues alone.

In order to test this prediction, Payne and Hauty (**5**) trained *Ss* to a common criterion on paired associates presented always in the same serial order. Upon completion of criterial trials, half the *Ss* received placebos while the rest received 25 mgm. doses of promethazine. Following a retention period of 75 min., half the *Ss* in each medication group underwent relearning trials wherein the paired associates were presented in the same order as before. The remaining *Ss*, exposed to randomized orders on each relearning trial, were deprived of secondary cues arising from serial positions and preceding responses. Retention efficiency was significantly lower for both medication groups under the random series condition than under the fixed series condition, but the deficit was much larger under the promethazine condition than under the placebo condition. Thus, the second prediction was confirmed.

The present paper reports a further test of the notion that the effects of depressant drugs upon learned behavior can be explained in terms of the hypothetical variables *reaction potential* and *reaction threshold*. From what has gone before it follows that additional training on the paired associate task prior to cue reduction should increase the reaction potential associated with the original cue compound, hence also the potential associated with what will become residual cues during relearning, so that the negative influence of augmented threshold *and* cue reduction will be at least partially offset. This prediction was tested in a modified replication of the paired-associate experiment wherein half the *Ss* in each combination of medication and relearning procedure met a criterion of 0% overlearning during the original learning session, while the rest met a criterion of 50% overlearning, as defined by Luh (**3**).

Procedure

Subjects.—The *Ss* were 128 volunteer basic airmen.

Task materials.—The 10 pairs of 2-syllable adjectives having the lowest average similarity ratings were selected from the list developed by Melton and Safier (see **1**, Table 11). These were typed in random order on the tape of an exposure device for use in the original learning trials. Additional random orders were derived and taped for use as required in the relearning trials. Multiple sets of exposure windows were manipulated in combination with multiple positions of lists on the tapes in order to provide sufficient variations of order.

Experimental operations.—The principal independent variables of the experiment were (*a*) degree of learning (0% vs. 50% overlearning), (*b*) order of paired associates during relearning (originally learned order vs. revised order on each relearning trial), and (*c*) drugs (placebo vs. 25 mgm. promethazine).

The *Ss* were assigned independently and randomly to the eight combinations of conditions, with the usual restriction imposed by equal subclass *N*'s. All were then trained on the fixed-order list by the method of anticipation to a criterion of two successive correct repetitions. Initial training was terminated at this point for those falling in the 0% overlearning group, but was carried on without interruption until the remaining *Ss* had satisfied the 50% criterion. Details of stimulus presentation were as follows: The cue term was exposed alone for 3 sec. during which *S* responded by pronouncing its associated term; then, both terms of the pair were exposed for 3 sec; then, the next cue term alone, and so on through the list. Twelve seconds intervened between the end of the exposure period for the tenth pair and the exposure of the first cue term alone. Exposures were controlled by a constant speed drive applied to the drum through a Geneva gear assembly.

Immediately following satisfaction of the overlearning criteria, all *Ss* were given appropriate medications orally in No. 1 pink capsules. Seventy-five minutes after medication, half the *Ss* in each combination of overlearning and

medication underwent relearning trials wherein the paired associates were presented in the same order as before, while the rest were exposed to revised orders on each relearning trial. Relearning was terminated upon completion of two successive errorless trials.[1]

Results and Discussion

Experimental outcomes were appraised in terms of trials required to relearn to criterion, although the savings score would have served as well as an index of retention efficiency. Relearning measures were transformed into common logarithms in order to circumvent inequalities among subclass variances, then analyzed by covariance methods using trials to initial criterion (excluding "overlearning" trials) as the independent variable. This analysis is shown in Table 1, and the subclass means are given in Table 2.

TABLE 1

Analysis of Adjusted Variation of Transformed Relearning Scores

Source	*df*	*MS*	*F*
Drugs	1	1.1369	33.74**
Learning	1	.2195	6.51**
Relearning	1	2.6084	77.40**
D × L	1	.1099	3.26
D × R	1	.1992	5.91**
L × R	1	.0002	<1.00
D × L × R	1	.0067	<1.00
Error	119*	.0337	
Total	126*		

* Reduced by 1 *df* for regression of transformed relearning trials upon original learning trials (excluding overlearning trials); $r_{xy} = .53$; $b = .0294$.
** $P \leq .02$. For D × L ratio, $P = .08$.

TABLE 2

Adjusted Means of Transformed Relearning Scores

Drugs	0% Overlearning		50% Overlearning		Means
	Fixed Order	Random Order	Fixed Order	Random Order	
Placebo	.46	.67	.45	.64	.55
Promethazine	.64	.99	.48	.86	.74
Means	.55	.83	.46	.75	
	.69		.61		.65

One can readily perceive that all three experimental variables have had significant effects upon relearning efficiency, but two of the primary interaction variances suggest that the drug effect is correlated, first, with the overlearning effect, and second, with the reduced cue effect. Upon closer examination of the first interaction it is apparent that the drug effect is much less pronounced under the "overlearned" condition than under the control condition, although double-tailed *t* tests show it significant under both conditions ($P \leq .01$). From the other viewpoint, the overlearning effect is more prominent under the promethazine condition ($P < .01$) than under the placebo condition (ns). Thus, the first of the previous experiments (**4**) stands confirmed in all essential respects, and one can have added confidence in the hypothesis that overlearning attenuates the effects of depressant drugs upon learned behavior by raising reaction potential sufficiently above the elevated threshold.

Turning now to the D × R interaction, the drug effect is much more

[1] Experimental data for this study were collected while the author was located at the School of Aviation Medicine, USAF. The pharmaceuticals were prepared by the Department of Pharmacology and Biochemistry under the direction of L. J. Milch, and medical safety aspects were monitored by W. L. Shepard and J. R. Kent of the Department of Aviation Medicine. Indebtedness is acknowledged to Samuel H. Mann for his able assistance in the collection of data.

pronounced under the random series condition than under the fixed series condition, although it is significant under both conditions ($P \leqq .02$). From the other viewpoint, the reduced cue effect is more prominent under the promethazine condition than under the placebo condition, although it is significant under both conditions ($P < .01$). Thus, the predicted effect of operations designed to reduce reaction potential, and the predicted behavior of this effect in relation to threshold manipulation, stand confirmed once again (**5**).

The proposition that overlearning will attenuate the negative effect of reduced cues (through its positive effect upon the reaction potential associated with residual cues) requires a significant L × R interaction in the present experiment. Moreover, the proposition that overlearning will mitigate the joint effects of drugs and cue reduction requires a significant second-order interaction. Obviously, neither of these requirements has been met. The overlearning effect is evident enough; but it occurs with about equal prominence under both fixed and random order conditions, and a detailed examination of it within the four D × R subclasses shows it confined almost wholly to the promethazine groups. The subclass distribution of raw means, shown in Table 3, favors the hypotheses of L × R and D × L × R interaction, but not enough to warrant rejection of null alternatives.

It is considered likely that higher degrees of overlearning might have produced results more consistent with theory. Under the combined effects of drugs and reduced cues, reaction evocation can be regarded as having been placed in a kind of "double jeopardy"—from elevated thresholds, on the one hand, and lowered reaction potentials, on the other. In fact, these two effects seem related synergistically, their joint effect being 1.36 times the sum of their separate effects under the minimal learning condition. One might argue, therefore, that 50% overlearning was barely adequate to compensate the drug effect, let alone remaining effects. Another possibility, hard to evaluate, is that boredom generated during overlearning trials may have attenuated the motivational component of the ${}_sE_R = f({}_sH_R) \times f(D)$ relationship during relearning trials. In view of overlearning effects obtained, however, this second possibility seems more remote than the first. Further experimentation should resolve the question.

TABLE 3

RAW MEANS OF RELEARNING SCORES

Drugs	0% Overlearnnig		50% Overlearning		Means
	Fixed Order	Random Order	Fixed Order	Random Order	
Placebo	3.75	4.87	3.06	4.62	4.08
Promethazine	4.56	11.62	4.31	8.94	7.36
Means	4.16	8.52	3.69	6.78	
	6.20		5.23		5.72

SUMMARY

This paper summarizes three experiments purporting to show how the effects of depressant drugs upon learned behavior can be explained in terms of the Hullian constructs *reaction potential* and *reaction threshold*. Predicted response changes resulting from the manipulation of habit strength and drive variables under drug and placebo conditions were confirmed in the first two experiments, but operations employed in the third experiment failed to elevate habit strength sufficiently to overcome the joint barrier imposed upon response evocation by drug-induced thresholds and lowered reaction potentials. However, the third experiment provided remarkable confirmation of the first two experiments.

References

1. Hilgard, E. R. Methods and procedures in the study of learning. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
2. Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.
3. Luh, C. W. The conditions of retention. *Psychol. Monogr.*, 1922, **31**, No. 3 (Whole No. 142).
4. Payne, R. B. Some psychological factors governing the impairment of verbal retention by cerebral depressants. *USAF Sch. Aviat. Med. Rep.*, No. 55–52, 1955.
5. Payne, R. B., & Hauty, G. T. Some psychological factors governing the effects of cerebral depressants upon learned behavior. *Arch. Int. Pharmacodyn.*, 1957, **111**, 470–477.

(Received April 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

CONDITIONED-STIMULUS VARIABLES IN AVOIDANCE LEARNING

MARVIN SCHWARTZ

U. S. Naval School of Aviation Medicine[1]

This report presents data showing the effects of varying several parameters of the conditioned stimulus (CS) in an avoidance learning procedure. Kessen (**1**) has shown that response strength in an avoidance procedure is a function of CS intensity. Kish's work (**2**) indicates that an on-going light is a more effective CS for avoidance learning than an off-going light. There are, however, other parameters of the CS which might be varied also and, in addition, there are no data presently available showing how these various parameters interact with each other. Accordingly, the present experiment tested the effects of varying the direction of change of CS energy, the rate at which the change was effected, the absolute amount of the change, and the CS-UCS interval.

METHOD

Apparatus.—The basic apparatus was a shuttle box that was divided into two identical compartments, each 15 in. long, 9 in. high, and 6 in. wide. The walls and ceiling of the box were made of clear Plexiglas. The floor of each compartment consisted of ¼-in. brazing rods, ½ in. apart, in a counterbalanced Plexiglas frame, to form two separately chargeable grids. A piece of ¼-in. plywood, painted flat black, extending from the ceiling to just above the grid floors, separated the compartments. A 3½-in. square hole, placed 2 in. above the grid floors, permitted access between compartments. The weight of a rat on one of the grids actuated a microswitch placed under it. One side of the assembly was sealed to a plywood tunnel 45 in. long and of appropriate cross section. The near end of the tunnel was faced with a piece of translucent frosted glass. The far end contained sockets for the insertion of the stimulus lights. The entire inside surface of the tunnel was painted flat white. When the stimulus lights were operating, an even light was diffused throughout both of the experimental compartments.

Illumination level was controlled by varying the voltage output of a Variac. Two microswitches were keyed to the arm of the Variac so that they were actuated only at the 0 − v. and 115-v. output points of the Variac. These microswitches were part of the shock circuit so that shock could be delivered to *S* when the illumination was either at its lowest or highest level. Two sources of illumination were employed. A 60-w. frosted bulb operating at 115 v. provided 25 ft.-candles in each compartment. A 7.5-w. bulb at the same voltage provided 3 ft.-candles. When the lights were off, the residual illumination was not measurable with the Norwood Director Incident Light Meter used to obtain the previous readings and barely permitted *E* to locate *S* in the apparatus.

A change in the illumination was used as the CS. After an appropriate interval, the grid was charged. The response required of *S* to escape the shock was crossing from the compartment he was then in to the opposite compartment. This response also returned the illumination to its original level and stopped a 1/10-sec. Standard Electric timer, providing a measure of latency of response. If the response was made between the time of change in illumination and the onset of the shock, the timer stopped and the illumination returned to its original condition and the shock was avoided.

The shocker consisted of a transformer with a 350-v. secondary and a 680,000-ohm resistor in series with the grids.

The experiment was conducted in a windowless, soundproofed, air conditioned room.

Subjects.—The *S*s were 32 male albino rats, 80–110 days old, purchased from the Budd Mountain Breeding Farm, Red Bank, N. J. Two *S*s were assigned at random to each of the 16 experimental conditions and the order in

[1] Opinions and conclusions expressed in this paper are those of the writer, and not necessarily those of the Navy. The writer wishes to express his appreciation to Wilse B. Webb, Director of the Aviation Psychology Laboratory, for his constructive criticism during the writing of this report.

which each *S* was trained was also assigned at random. For three days prior to training, each was handled for 3–5 min. daily. During this time, *Ss*, in numbers up to three at a time, were allowed to explore the experimental apparatus for 10–15 min. daily. They were housed two to a cage in a different building and given food (Purina pellets) and water ad libitum. They were carried by car, in a carrying cage, to the experimental room.

During the experiment, 7 *Ss* were replaced from the same stock; two for learning to straddle the barrier between compartments of the apparatus, two for refusing to escape the shock, and three for shuttling between compartments continuously without letup.

Experimental design.—Four parameters of the CS were varied in the present experiment: (*a*) direction of illumination change, i.e., light On vs. Off; (*b*) rate of illumination change, i.e., light-change essentially instantaneous owing to the operation of a knifeswitch (Fast) vs. carried out over the entire CS-UCS interval by gradually changing the voltage output of the Variac (Slow); (*c*) amount of change in illumination, i.e., the 7.5-w. (Low) vs. the 60-w. bulb (High); and (*d*) the CS-UCS interval, i.e., 3 sec. vs. 6 sec. It should be noted that the Slow illumination change was accomplished by manually varying the voltage output of the Variac between 0 and 115 v. in as smooth a fashion as possible. As previously described, shock switches were keyed to these values of the Variac and, hence, there was some variability in the CS-UCS interval. Shock was delivered at some time between 2.5 and 3.0 sec. in the shorter interval and 5.5 and 6.0 sec. in the longer interval. The combination of these four variables produced a factorial design which enabled assessment of the influence of each variable independently as well as their several interactions.

Procedure.—The entire experimental procedure for each *S* was carried out in a single session. At the start of the session, *S* was placed in the apparatus, with the illumination in the between-trials condition, and allowed 5–10 min. habituation. Immediately following this, avoidance training was begun. Fifty trials were administered, spaced a minimum of 50 sec. apart, with the stipulation that there be no spontaneous crossing between compartments during the final 10 sec. of the intertrial interval. Immediately following Trial 50, the shock circuit was turned off and extinction trials were run in the same manner as training trials. The criteria of extinction were five consecutive trials in which no response occurred within 30 sec. following the illumination change, or a maximum of 25 trials. Although two CS-UCS intervals were employed during training, during extinction the criterion for avoidance was a response in 6.0 sec. or less for all *Ss* in order to allow all *Ss* an equal opportunity to demonstrate the avoidance response.

If *S* did not receive a shock on Trial 50 because he successfully avoided, that trial and all previous successive avoidances back to the last trial on which *S* did receive shock were scored as having occurred during extinction. Therefore, there existed the possibility of exceeding 25 extinction trials. This criterion was adopted because an *S* who avoided on Trial 50, and on successive trials previous to this trial, received no further primary reinforcement and was actually in the process of extinction.

Results

Table 1 presents the mean total number of avoidances during the 50 acquisition trials as a function of each of the four variables manipulated in the present experiment. As can be inferred from these data, the overall response level was low. That learning did occur, however, is indicated by the fact that only 3 of 32 *Ss* avoided the shock, once each, during the first 10 training trials while 20 of 32 avoided at least once during the final 10 trials. "Spontaneous" responses between trials were exceedingly rare.

Table 2 presents the summary of the analysis of variance performed on the total number of avoidances during acquisition. Although the variances were not homogeneous, the discrepancy was not great enough to invalidate the use of the analysis (**3,**

TABLE 1

Mean Total Number of Avoidance Responses During Acquisition for Each of the Four Main Variables

(N = 16 for all comparisons)

On = 11.94	Low = 5.62
Off = 6.56	High = 12.88
Fast = 8.50	3 sec. = 4.31
Slow = 10.00	6 sec. = 14.19

TABLE 2

ANALYSIS OF VARIANCE OF TOTAL NUMBER OF AVOIDANCE RESPONSES DURING ACQUISITION

Source	*df*	*MS*	*F*
On-Off (O)	1	231.13	6.40*
Fast-Slow (F)	1	18.00	
High-Low (H)	1	420.50	11.64**
CS-UCS Interval (C)	1	780.13	21.59**
O × F	1	21.12	
O × C	1	31.99	
O × H	1	15.12	
F × C	1	45.12	
F × H	1	220.50	6.10*
C × H	1	36.12	
O × F × C	1	31.98	
O × F × H	1	55.10	
O × C × H	1	127.98	
F × C × H	1	15.13	
O × F × C × H	1	2.03	
Residual	16	36.13	

* $P < .05$.
** $P < .01$.

p. 78). As in Kish's experiments (**2**), the On-Off difference proved to be significant beyond the .05 level ($F = 6.40$; 1 and 16 *df*). The On group was superior to the Off group, but unlike Kish's *Ss*, the Off group did demonstrate appreciable learning, some *Ss* performing as well as any in the On group. Of the 16 *Ss* in the Off group, 7 gave a minimum of 4 avoidances.

Illumination level (High–Low) and CS-UCS interval differences were each significant beyond the .01 level. The *Ss* in the High illumination conditions performed better than those in the Low conditions and *Ss* in the 6-sec. CS-UCS interval group were superior to those in the 3-sec. group. In addition, the Fast–Slow and High–Low interaction was significant beyond the .05 level. Performance was superior in the Low conditions if the change was Fast and in the High conditions if the change was Slow. The respective means for Low and High conditions under a Fast change were 7.50 vs. 9.50 and 3.75 vs. 16.25 for Low and High conditions under a Slow change. The Fast–Slow difference did not prove to be significant.

Responses during extinction were also analyzed. There were only two major differences from the acquisition results: (*a*) The On–Off difference did not prove significant during extinction ($F = 2.57$; 1 and 16 *df*) and, (*b*) The On–Off and Fast–Slow interaction was significant beyond the .01 level ($F = 9.02$; 1 and 16 *df*). The respective means for the On and Off conditions under a Fast change were 6.00 vs. 9.25 and 13.25 vs. 2.95 under the Slow change condition. Thus, *Ss* in the On condition performed better during extinction if the change in illumination was Slow while *Ss* in the Off condition performed better if the change was Fast. It should also be noted that over 25% of the *Ss*, about equally divided between the On and Off conditions, did not reach extinction in the allotted 25 trials.

DISCUSSION

The results of the present study show that during acquisition performance was superior in those groups having the ongoing light as the CS for avoidance. This difference did not, however, hold up during the extinction process. In addition, it was noted that during acquisition many *Ss* in the Off group performed as well as any in the On group. These findings differ from Kish's inasmuch as he states that in two of his experiments learning in the Off conditions "was so small as to leave this issue in doubt" (**2**, p. 35).

It is hypothesized that the clearer demonstration of learning in the Off condition in this experiment when compared with Kish's results can be accounted for in terms of the response required of *S* or the training conditions.

A demonstration experiment was performed which tends to support this hypothesis. The

conditions were comparable to Kish's with two exceptions: (*a*) Since the tendency to crouch in the presence of electric shock and/or light is strong in albino rats, and, further, since this crouching seriously interferes with the performance of a wheel-turning response, this tendency was eliminated by only utilizing animals as *S*s if they did not display this response in a series of three preliminary, shock-escape trials given prior to the avoidance training. (*b*) Training was spaced over a 20-day period, 5 trials per day, instead of all in a single session. In all other respects, the procedure was comparable to Kish's, i.e., a wheel-turning response with a 6-sec. CS-UCS interval and a 30 ft.-candle change in illumination were employed. Five *S*s were trained to avoid a shock by turning a wheel. The CS was an off-going light. After 100 training trials, the *S*s were avoiding approximately 90% of the time. In terms of both probability of response and response latency, acquisition was significant beyond the .01 level. Thus, an efficient avoidance performance can be established utilizing an off-going light as the CS.

From this latter experiment and the difference between the results obtained in the main experiment and Kish's, it would appear that the experimental conditions under which training is conducted are a more important determinant of avoidance performance than is the direction of change of CS energy.

The present experiment showed that during both acquisition and extinction, performance was superior in those groups having the greater change in CS energy. This effect was independent of the direction of change in energy and the rate at which the change was effected. In fact, the rates of change employed had no discernible independent effect. These results are consistent with Kessen's finding (**1**) that performance during acquisition is a function of CS intensity where the CS is an on-going light.

Contrary to expectation, the present experiment showed that avoidance performance was superior where the CS-UCS interval was 6 sec. as compared to 3 sec. This is also inconsistent with Kish's data. The discrepancy would appear to be ascribable to the nature of the respective responses required of *S*. Typical rat behavior in a wheel-turning situation, under massed acquisition, such as Kish employed, consists of "hanging on" the wheel and rarely moving away from it. Also, a high output of "spontaneous" responses characterizes such a procedure. In contrast, in the present shuttlebox, all *S*s invariably stayed at the extreme ends of the respective compartments and rarely made spontaneous crossings. It appears likely, therefore, that the present results can be accounted for on the assumption that the 3-sec. interval was simply too short for *S* to perceive the CS change and complete the response in time to avoid the shock. Thus, they were probably shocked while in the act of performing the response. If this occurred with any consistency, such punishment of the avoidance response would be expected to disrupt the avoidance pattern. The observation of abortive avoidance responses in the 3-sec. group lends credence to this hypothesis. On the other hand, the 6-sec. group had sufficient time in which to complete the entire process successfully before the shock came on. In a recent review, Solomon and Brush come to similar conclusions when comparing avoidance with a 2.5-sec. and a 5.0-sec. CS-UCS interval (**4**, p. 260).

Finally, there were strong indications in the present experiment that the parameters manipulated interact with each other. During acquisition, the rate of CS change interacted significantly with the absolute amount of CS change and during extinction it interacted significantly with the direction of CS change. While further work is in progress to clarify the nature of these interactions, it is tentatively hypothesized that they influence the discriminability of the CS.

Summary

The present report presents data concerned with the effects of varying several parameters of the CS on performance of an avoidance response. The following conclusions were reached: (*a*) Avoidance response strength is greater if the change in CS energy is greater,

irrespective of the rate or direction of the change. (*b*) The response required of *S* and the training procedure are probably of greater importance in determining performance differences than are differences due to the direction of change in CS energy. (*c*) In a shuttlebox, avoidance response strength is greater with a 6-sec. CS-UCS interval than with a 3-sec. interval.

References

1. Kessen, W. Response strength and conditioned stimulus intensity. *J. exp. Psychol.*, 1953, **45**, 82–86.
2. Kish, G. B. Avoidance learning to the onset and cessation of conditioned stimulus energy. *J. exp. Psychol.*, 1955, **50**, 31–38.
3. Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
4. Solomon, R. L., & Brush, E. S. Experimentally derived conceptions of anxiety and aversion. In M. R. Jones (Ed.), *Nebraska symposium on motivation 1956.* Lincoln: Nebraska Univer. Press, 1956. Pp. 212–305.

(Received April 22, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

AGE DIFFERENCES IN SHORT-TERM RETENTION OF RAPIDLY CHANGING INFORMATION

WAYNE K. KIRCHNER[1]

Minnesota Mining and Manufacturing Company

There is need in the field of aging for an experiment directly measuring very short-term retention. While older people tend to do well on memory span experiments, their performance on tasks demanding the acquisition and retention of continuously changing information is suspect. In industry, in particular, older people are found in small numbers on "paced" tasks (**1**) which seem to involve this type of performance.

Simply stated, the purpose of the present experiment was to measure the performance of older and younger *Ss* on a "paced" task involving rapidly, continuously changing information. The assumption is that on this type of task *Ss* must utilize short-term retention and that an analysis of the results should throw some light on such retention and its relationship to age.

METHOD

Apparatus.—The apparatus has been decribed fully by Welford (**3**). Primarily it is a visual display consisting of 12 Morse keys, spaced 2 in. apart and numbered from 1 to 12 from left to right, above which are 12 small lights, spaced with a light directly above each key and also numbered from 1 to 12.

By means of an automic programmer, each of the 12 lights can be made to come on in any desired sequence. The time interval between the disappearance of one light and the appearance of another is, for practical purposes, simultaneous and the duration of each light may be varied, although all lights in a series must have the same time duration. Also, any one of the 12 keys can be made to be the correct response for each position of the light. Further, the lights and keys in operation may be limited either to half of the keyboard (6 keys) or to the full keyboard (12 keys).

All responses (key pressings) made by each *S* were recorded and timed by means of an electronic recording device—the SETAR (**4**)—which also automatically determined the correctness of each response and punched all information, in binary code, onto paper tape.

Subjects.—Forty *Ss* were used. Twenty (7 men and 13 women) were older, ranging in age from 60–84 yr.; 20 (13 men and 7 women) were younger, ranging in age from 18–24 yr. The older *Ss* were obtained through local Red Cross clubs and an old people's home. No claim is made here that they were "typical" *Ss* as they were all volunteers who came from diverse occupations. The younger *Ss* were 10 sailors from the Royal Navy and 10 Cambridge University students from the laboratory practical class in psychology. The sailors were non-volunteers; the students were all volunteers.

The task.—Each *S* had to respond to a light which moved continuously from one position to another at a fixed rate of speed. Nothing that *S* did could change that rate of speed. For each position of the light, *S* had to press a key in relation to either where the light was or had been. If he were correct, a buzzer sounded; if not, silence, and he was at liberty to press other keys if he wished and/or if he had time.

Basically, there are four conditions defined as follows: (*a*) *No-back.* The *S* was instructed to press the key where the light appeared. The major function of this condition was to determine *S*'s ability to keep up physically with the moving light. (*b*) *One-back.* The *S* was instructed to press the key where the light had just gone out. He was following the light but was "one light behind" or "one-back." (*c*) *Two-back.* The *S* was instructed to press the key where the light had gone out two positions before. (*d*) *Three-back.* The *S* was instructed to press the key where the light had gone out three positions before.

[1] This study was conducted under a Fulbright research grant at Cambridge University, England in 1954–55, while the author was attached to the Nuffield Research Unit into the Problems of Aging, Cambridge. His sincere thanks go to A. T. Welford, Director of the Unit, for his advice, patience, and friendship.

The rate of speed in all conditions was either 1.5 sec. or 4.5 sec. for the duration of each light. The longer time interval was used only on Two-back for some selected older *Ss* and on Three-back for all younger *Ss* and the few older *Ss* able to attempt this condition. For nearly all trials, then, the light moved from one position to another every 1.5 sec., a time considerably longer than normal reaction time to a visual stimulus yet fast enough, it was hoped, to impose a load or stress on *S*. Indeed, as *S* progressed from condition to condition, the load became greater and greater as more information had to be perceived and retained in the same amount of time.

For all conditions, the lights came on in a quasi-random order, varying from trial to trial. Each sequence of lights was repeated three times on any one trial.

Considerable difficulty was found in explaining conditions beyond No-back to many of the older *Ss*. For that reason, the original rigid instructions were tempered and flexibility became the key. Still, lack of comprehension was apparent in some *Ss*.

Design.—An order of trials as listed in Table 1 was presented at the 1.5-sec. rate to each *S*. The *S* went as far as he could go on the various conditions. In terms of results, this meant that all of the younger group was able to attempt Three-back. Few of the older *Ss* could do Two-back well, however, so most stopped at that point.

In addition to the trials listed in Table 1, all younger *Ss* did two trials of 18 lights each (6 keys) and one of 36 lights (12 keys) on Three-back at the 4.5-sec. rate. A small group of older *Ss* did one trial of 36 lights (12 keys) on Two-back, also with the 4.5-sec. rate.

The experimental design is inadequate for the evaluation of the effects of using 6 or 12 keys. Therefore, the 6-key conditions were used as a "warmup," and were especially valuable for this purpose for the older *Ss*. All major comparisons are made with the 12-key results.

TABLE 1

ORDER OF CONDITIONS AND DESIGN OF EACH

Condition	Trials	Lights/ Trial	Possible Correct Responses
1. No-back (6 keys)	2	18	36
2. No-back (12 keys)	1	36	36
3. One-back (6 keys)	2	18	36
4. One-back (12 keys)	1	36	36
5. Two-back (6 keys)	2	18	36
6. Two-back (12 keys)	1	36	36
7. Three-back (6 keys)	2	18	36
8. Three-back (12 keys)	1	36	36

Analysis of data.—Older and younger *Ss* were compared in terms of the following scores: (*a*) number (and percentage) of responses correct, whether or not the correct response was the first response made; (*b*) total number of responses; and (*c*) the number and type of errors. Five types of errors were identified: (*a*) *Omission.* The *S* makes no response (key press) to a light. (*b*) *One Key Early.* The *S* presses a key one ahead of correct key in over-all serial order. For example, correct order of keys is 7, 5, 10 and *S* presses Key 5 when he should be pressing Key 7. (*c*) *Spatial.* The *S* presses a key spatially adjacent to correct key. (*d*) *One Key Late.* The *S* presses a key one behind correct key in over-all serial order, i.e., in series 7, 5, 10, *S* presses Key 5 when he should be pressing Key 10. (*e*) *Random.* All other errors are considered as random ones.

RESULTS

In general, the older *Ss* were unable to perform as well as the younger *Ss*. Only three of the older group were able to reach the Three-back stage. For this reason, all comparisons involving older *Ss* were made through the Two-back stage only.

Absolute and relative differences.—Table 2 presents the mean number of correct responses and errors made by each group under the various experimental conditions. This table shows rather conclusively that the performance of older *Ss*, in absolute terms, was inferior to that of younger *Ss*. Older *Ss* made fewer correct responses, a smaller total number of responses, and more errors, when One-back and Two-back conditions are considered. Their performance decreased markedly as the condition was changed from One-back to Two-back. In addition, older *Ss* showed greater variance in all phases of the results, indicating a much greater range of performance.

An analysis of each condition reveals the following: (*a*) For the No-back condition, young and old

TABLE 2

MEAN NUMBER OF CORRECT RESPONSES, ERRORS, AND TOTAL RESPONSES MADE BY OLDER AND YOUNGER GROUPS*

Group	Condition											
	No-back			One-back			Two-back			Three-back		
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*
Total Correct Responses												
Sailors	9	35.89	.14	10	35.70	.64	9	30.00	5.01	10	16.90	5.86
Students	9	36.00	.00	9	36.00	.00	10	32.90	4.13	10	24.00	7.38
Old people	16	35.31	1.77	18	29.33	11.05	12	12.58	6.78	—	—	—
Total Errors												
Sailors	9	.11	.32	10	.30	.64	9	7.56	5.53	10	19.40	6.76
Students	9	.00	.00	9	.00	.00	10	3.50	4.72	10	13.90	6.16
Old people	16	.88	1.69	18	6.89	11.14	12	24.08	6.21	—	—	—
Total Number of Responses												
Sailors	9	35.89	.14	10	35.90	.30	9	34.89	2.67	10	30.70	4.90
Students	9	36.00	.00	9	36.44	1.11	10	35.40	2.06	10	34.40	5.69
Old people	16	35.63	1.28	18	33.61	5.08	12	24.25	7.96	—	—	—

* Total errors for an *S* could exceed the number of lights per trial, since more than one error could be made for each light.

were approximately equal in their ability to keep up with the simple physical movement of the light. No significant differences were observed. (*b*) On the One-back condition, younger *Ss* performed as well as on No-back. Older *Ss*, however, dropped off in performance with mean differences of over six points in correct responses, over two points in total responses, and over six points in errors being found between young and old. (*c*) The Two-back condition gave the greatest mean differences. Older *Ss* made fewer correct responses (over 17 points), fewer total responses (over 10 points), and more errors (over 17 points) than the worst group of younger *Ss* (sailors) on this condition. All of these differences are highly significant statistically.

Clearly, in terms of absolute results, the older group lagged far behind the younger groups. These results, too, are based on those older *Ss* who were able to attempt the various conditions. On Two-back, for example, eight *Ss* were not included because they either failed to understand what was required or gave up entirely after a few lights and thus were classified as "failures." Obviously, if these *Ss* had been included, the older averages would have suffered still more.

The table further indicates that the performance of younger *Ss* on Three-back approached that of the older *Ss* on Two-back although the younger

TABLE 3

PERCENTAGE OF TOTAL ACTUAL RESPONSES CORRECT FOR OLDER AND YOUNGER GROUPS ON ALL CONDITIONS

Group	Condition			
	No-back	One-back	Two-back	Three-back
Sailors	100.00	99.4	86.0	55.0
Students	100.00	98.8	92.9	69.8
Old people	99.1	87.3	51.9	—

group still made more correct responses and fewer errors. Likewise, the performance of younger *Ss* on Two-back was roughly equivalent to that of older *Ss* on One-back. If an analogy of climbing a stepladder can be used here, it suggests that older *Ss* tend to fall off one "rung" before the younger *Ss* do. Presumably, both perform as well "on the ground" (No-back).

These results, of course, are measured in absolute terms. What of the performance of older persons in relative terms? Were the older *Ss*, as so often appears to happen in the everyday world, able to make fewer responses in all, but able to make a high proportion of those responses good ones? Table 3 presents the answer from this experiment.

Once again, the fact emerges that the younger group exceeded the older group in performance, this time in relative terms. If these figures can be thought of as a measure of efficiency, then older people clearly are not as efficient. For example, on Two-back, only about half of the key pressings made by the older group were correct while almost 90% of those of the most inefficient younger group (sailors) were correct.

Type of error.—An analysis was made of the different types of errors made by each group and the relative frequency of their occurrence. The Two-back condition was chosen for detailed comparison of the young and old because the younger *Ss* made few, if any, errors on One-back and no valid comparison could be made on that condition with the older *Ss*' records.

Table 4 presents the outcome of this analysis. Included, too, in the table is the record of the older *Ss* on One-back and of the younger *Ss* on Three-back.

Major differences between young and old on Two-back are these:

TABLE 4

TYPE OF ERROR MADE BY OLDER AND YOUNGER *Ss* UNDER VARIOUS CONDITIONS

Group	Condition	Type of Error										Total Errors
		Omission		One Key Early		Spatial		Random		One Key Late		
		N	%	*N*	%	*N*	%	*N*	%	*N*	%	
Old people	1-back	49	39.5	8	6.5	46	37.1	16	12.9	5	4.0	124
Sailors	2-back	24	35.3	14	20.6	15	22.1	13	19.1	2	2.9	68
Students	2-back	13	37.1	7	20.0	7	20.0	5	14.3	3	8.6	35
Old people	2-back	153	52.9	36	12.5	24	8.3	71	24.6	5	1.7	289
Sailors	3-back	67	34.5	22	11.3	26	13.4	74	38.1	5	2.6	194
Students	3-back	45	32.4	39	28.1	23	16.5	28	20.1	4	2.9	139

TABLE 5

COMPARISON OF OLDER AND YOUNGER *Ss* ON THREE-BACK AND TWO-BACK CONDITIONS WHEN ALLOWED MORE TIME (4.5-SEC. RATE)

Group	Correct Responses			Total Responses			Errors			% of Total Responses Correct
	N	Mean	*SD*	*N*	Mean	*SD*	*N*	Mean	*SD*	
Sailors (3-Back)	10	31.20	4.38	10	37.00	3.46	10	7.40	4.54	84.3
Students (3-Back)	10	34.50	7.68	10	37.20	2.04	10	1.90	4.13	92.7
Old people (2-Back)	4	25.25	9.83	4	29.75	7.14	4	13.00	12.53	84.9

Older *Ss* made a greater percentage of Omission and Random errors and a smaller percentage of One Key Early and Spatial errors. Taken together, these results indicate that older *Ss* could not keep up as well (Omission), were "lost" more often (Random and Spatial), and did not "get ahead of the game" (One Key Early).

On One-back, there is a different picture. Here, older *Ss* closely resembled the younger *Ss*' performance on Two-back with the relative amount of Omissions about the same. Older *Ss* on One-back made larger numbers of Spatial errors, however.

Results are not as clear when the errors of younger *Ss* on Three-back are compared with the Two-back performance of the older group. However, the older group still maintains a lead in terms of Omission errors.

A general picture of the average older *S* is fairly clear. His performance on No-back was about the same as the younger *Ss*. He fell off on One-back with his major errors being Spatial ones, indicating that he had the right idea of which key to press although suffering in accuracy. On Two-back, a great decrease in performance occurred with the older *S* sitting before the display and failing to respond to many of the lights; and when responding, making many Random errors.

The younger *S* appeared to do equally well on No-back and One-back with his first decrease in performance occurring on Two-back followed by a greater drop on Three-back. His record on these two tasks was one of inaccurate responding, not one of Omission, as characterized the older *S*'s major failing.

Effect of more time.—The above results readily lead to a hypothesis that the time factor plays an important part in the results and that older *Ss*, if given more time, might do much better and approach or surpass the younger group in performance. As a test of this hypothesis, four older *Ss*, who were unable to do the Two-back condition with any skill when the light moved at 1.5-sec. intervals, were allowed to do the task with the light moving at 4.5-sec. intervals.

Results for this small group of *Ss* are shown in Table 5 along with similar results from younger *Ss* who did the Three-back task at this slower rate.

The data in this table and Table 2 reveal that these four *Ss*, on the average, made over 12 more correct responses, over 5 more total responses, and over 11 fewer errors than the older *Ss* who undertook Two-back with the regular speed of 1.5 sec.

Comparison between young and

old *Ss* on Table 5 reveals nothing new. Once more the younger *Ss* are appreciably better. Certainly, and not too surprisingly, the allowance of more time has a salutary effect on performance of both young and old.

Discussion

While it may be a secondary factor, the speed of movement of older persons is not the crucial factor in their poorer performance. This is shown by the equal performance of young and old on the No-back condition. Directly related to this are Singleton's findings (**2**) that times taken for actual *movements* by older *Ss* on a joy-stick were as fast or faster than those of younger *Ss*.

The crucial factor is contained somewhere in the fact that at a certain point—the Two-back condition—older *Ss* were unable to perform as well. At this point occurs a large number of omissions, indicating an inability to cope with the situation. If the motor performance of the older persons is not at fault, then the answer must lie elsewhere, and probably in what can be called the central organizing processes of the brain.

A short consideration of the Two-back conditions for a moment may serve to point out the nature of these central processes. The *S* on this task is storing two pieces of information somehow as he waits for a third piece of information to appear. This new piece of information will serve as a signal for his response to the first piece of information being held in storage. Of course, he must also put this new datum into storage before he can await further pieces of information to appear. A continual process of mental give-and-take is going on and all these incoming and outgoing streams must be organized somehow. The function of the central processes is to do just that.

If the answer is the central organizing processes, then evidently the poorer performance of the older *Ss* must be related to a "slowing down" of the central processes. That this is so is suggested by the results of the four older *Ss* who, when given more time, were able to change their performance on Two-back from one of almost complete failure to a much more satisfactory one. In relative terms, they approached the performance of the younger groups. Perhaps this is not too surprising as the task with more time approaches one of memory span for digits, where older *Ss* hold up well.

Over all, it does seem clear that as the time interval allowed for retention becomes shorter and shorter, older *Ss* will become more handicapped and will reach their point of handicap much sooner than younger *Ss*. In this experiment, where retention times were shortened by the simple expedient of forcing *S* to retain more and more data over the same time interval, the evidence points to one possible and probable explanation for the poorer performance of older *Ss*—a slowing down of the ability to organize the continuous interchange of incoming and outgoing information.

The bad effects of "paced" tasks in industry and everyday life perhaps can be traced to this difficulty in short-term retention. Certainly, further work on this type of retention must be done before all answers to the problem of aging can be given.

Summary

An experiment was conducted to measure very short-term retention in younger and older *Ss* by means of a visual display involving a rapidly moving light. The basic time interval between lights for all stages of the experiment was 1.5 sec.

Results indicated that older *Ss* slumped in performance much sooner than younger *Ss*, in both relative and absolute terms. Older *Ss* also tended to make more errors of omission and more random responses, indicating a lack of ability to "keep up." As little or no difference in physical ability to follow the moving light was revealed between young and old, it

was suggested that the inability to organize incoming and outgoing information as rapidly as the younger *Ss* caused the older *Ss*' poorer performance.

References

1. Belbin, R. M. Difficulties of older people in industry. *Occupational Psychol.*, 1953, **27,** 177–190.
2. Singleton, W. T. The change of movement timing with age. *Brit. J. Psychol.*, 1954, **45,** 166–172.
3. Welford, A. T. An apparatus for use in studying serial performance. *Amer. J. Psychol.*, 1952, **65,** 91–97.
4. Welford, N. T. An electronic digital recording machine—the SETAR. *J. Sci. Instrum.*, 1952, **29,** 1–4.

(Received April 23, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

A NEW TECHNIQUE FOR STUDYING SPATIAL GENERALIZATION WITH VOLUNTARY RESPONSES[1]

JUDSON S. BROWN, FRANK R. CLARKE[2], AND LARRY STEIN[3]

State University of Iowa

The term *spatial generalization* derives its meaning from studies which show that a response, after it has been associated with a given stimulus, tends to be elicited, though often with a decrement, by identical stimuli whose spatial position differs from that of the first. In studies such as those of Anrep (**1**) and of Bass and Hull (**2**), classical conditioning procedures were employed, the position of the generalized stimuli being varied along the surface of *S*'s skin. More recently, Gibson (**5**), and Brown, Bilodeau, and Baron (**4**) obtained analogous results in discrimination reaction-time tasks where *Ss* were instructed to respond to one stimulus but not to others whose spatial location differed from that of the first. A phenomenon similar to the generalization of conditioned responses was observed in both experiments since the frequency of false reactions to the prohibited stimuli increased as a function of their nearness to the approved stimulus.

The generalization-like gradients obtained in these latter studies are reasonably stable and their presence has been confirmed in other investigations (**3, 6**). Nevertheless, the general utility of the method is limited because of certain overly critical procedural details. The *Ss* must be constantly urged to hurry if adequate numbers of false reactions are to be obtained; a strong set to react to the training stimulus must be maintained by frequent retraining trials to that stimulus; and the instructions must be phrased to condone implicitly, but not encourage, false reactions. Furthermore, the method cannot readily be applied to the study of the effects on generalized reactions of certain variables known to be significant in conditioned-response experiments. For example, it is difficult to vary the strength of the tendency to react to the approved stimulus. The description of an alternative method that seems to obviate most of these difficulties is the subject of this report.

Method

Apparatus.—The apparatus used with the new procedure was, in the main, like that described in detail before (**4**). Seven signal lamps (117 v., 7.5 w.) were arranged in a horizontal row at about eye level for a seated *S*. They were mounted on the surface of a black plywood panel which was curved so as to make them equidistant (5 ft.) from *S*'s face and were equally spaced at 8° intervals. Located directly below the middle lamp were two "win" and two "lose" signs that could be individually illuminated from behind. Directly in front of *S*, at table-top height, was his reaction switch. Pushing the lever of this switch forward (toward the stimulus panel) would cause one of the "win" signs to light up; moving the lever backward would light one of the "lose" signs. The *E* could light up either of the other two signs and any one of the seven signal lamps, according to a prearranged sequence.

Procedure.—The *Ss* were told they were participating in an experiment to determine "how well people can make predictions in a simulated horse-race betting situation." The seven stimulus lamps were said to represent seven different race horses and the lighting of a lamp meant that the symbolized horse was running a race against other imaginary horses

[1] Appreciation is expressed to Bettina Bass for a helpfully critical reading of the manuscript.

[2] Now at the University of Indiana.

[3] Now at VA Research Laboratories in Neuropsychiatry, Pittsburgh.

not represented by lamps. The *S*'s task was to guess whether the horse would win or lose that race and to express his guess by an appropriate movement of the reaction-switch lever. Following each guess, *E* flashed on either a "win" or a "lose" sign to provide *S* with information concerning the correctness or incorrectness of the guess. All "horses" ran an equal number of "races" during a given experimental session, their wins and losses being arbitrarily predetermined. The *S*s were told that their initial choices would have to be based on sheer guesses but that after a number of trials they should be able to get some idea as to how often each horse won and would thereupon be able to place their bets with increasing accuracy.

In a typical experiment, *S* was given 210 trials or "races" spaced at intervals of about 10–15 sec. These trials were divided into six blocks of 35 each. Within each block, every light was presented five times in randomized order. A different order was used for each block and the order of trial-block presentation was randomly varied from *S* to *S*. In the studies described below, the prearranged win-lose schedule was such that the horse symbolized by the middle lamp won 80% of its races, and all other horses won 20% of theirs, irrespective of *S*'s guesses. These objective win-lose percentages were maintained within each block of 35 trials by having the middle horse win a randomly selected four of five races and the other horses win one of five. In other studies, the empirical win percentages of the central and peripheral lamps have been varied considerably from these values.

In designing this procedure, it was anticipated that the frequently reinforced tendency to guess "win" to the middle lamp (80% winner) might generalize to the peripheral lamps in proportion to their nearness to the center, so as to produce bidirectional gradients of win-response frequency.

Illustrative Studies

Experiment I.—In this investigation, 15 men and 15 women from an elementary psychology class were tested according to the procedures outlined above. The results are summarized by the solid-line curve of Fig. 1. Here win responses made during the last five blocks of trials have been converted to percentages and have been plotted against the ordinal positions of the seven lights as arranged from left to right on the curved stimulus panel. The mean number of win responses can be obtained by dividing any percentage value by four. The empirical win percentages for each horse are given on the abscissa.

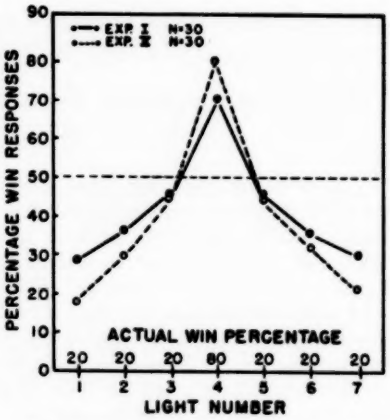


Fig. 1. Generalization-like gradients of win-response frequency obtained in two independent experiments under two sets of instructions.

As this figure shows, generalization-like gradients appear on each side of the 80% winner, lamps near the middle eliciting higher win-response percentages than more distant ones in spite of the fact that all lamps save No. 4 "won" with exactly the same empirical frequency (20%). The generalization effect is limited, of course, to the sections of the curves between Lights 1 and 3, and between 5 and 7. The steep drops from Light 4 to 3 and from 4 to 5 are not part of the generalization effect but an obvious consequence of the difference in actual win percentage. A statistical comparison of the means corresponding to Lights 1, 2, and 3 yielded a highly significant F ratio ($P = < .001$) as did a similar comparison among means for Lights 5, 6, and 7. The tentative conclusion is drawn, therefore, that the strong tendency to say "win" to the middle

light generalized to adjacent lights in proportion to their nearness to the center. The estimates of win-lose probabilities for the peripheral lamps seem to have been modified or distorted by the interposition of the "big winner."

The solid-line curve in Fig. 1 also shows that the estimates for all lamps differ considerably from the objective win-lose probabilities. The six peripheral lamps elicit win responses more often than 20% of the time and the value for Light 4 falls short of 80%. Because of this discrepancy the *Ss* were questioned concerning the ways in which they had tried to solve the problem. Their replies suggested that some of them had tried to "beat the game" by discovering a "correct" *serial sequence* of wins and losses. This incorrect "set" may have prevented them from paying adequate attention to the empirical win-lose frequencies. In the hope that problem-solving hypotheses of the serial type could be minimized, the instructions were rewritten for use in Exp. II.

Experiment II.—The procedural details of this study were identical with those of the first. The instructions were modified, however, to place even greater emphasis on the point that the outcomes of the races for the seven "horses" were completely independent of one another. The idea of independence was further stressed by urging *Ss* to think of the horses as racing on tracks in widely separated parts of the country. They were also told that their betting efficiency could be improved by trying to remember how often each horse won on its own track.

The dashed-line curve of Fig. 1 shows the results obtained from 30 undergraduate *Ss* (15 men, 15 women). Clearly the modified instructions had the desired effect of bringing all points nearer to the objective probabilities. The percentages at Lights 1, 4, and 7 approximate closely the empirical values of 20, 80, and 20. But it is also clear that the basic generalization gradients of Exp. I are again obtained, and are, in fact, even steeper than before.

Data of additional interest were obtained from questionnaires administered to 28 *Ss* of Exp. II following their experimental sessions. These *Ss* were asked to estimate (in percentage terms) the frequency with which each of the seven horses actually won. The averages of these estimated percentages, for Lamps 1 through 7, respectively, were 25, 31, 37, 80, 40, 29, 25. Again, the values for the peripheral lights form symmetrical generalization-like gradients, though they are slightly flatter than those obtained from the manual responses. Should one choose to regard the questionnaire data as indicating *S*'s "perceptions" of win likelihood, it follows that perceptions as well as motor responses suffer generalization-like modifications in this situation.

Discussion

Although bidirectional gradients were obtained in both of these illustrative studies, the possibility remains that they are due to factors other than generalization. For instance, the gradients might simply mirror preferences that *S* brings to the situation. To test this notion, tabulations were made of the win responses of 180 *Ss* (from Exp. I, II, and others) to the first presentation of each signal lamp. The percentages obtained for Lamps 1 through 7, were 54, 48, 58, 56, 51, and 54, respectively. From these figures, it is clear that the middle lamp is not the "most preferred" as a winner, that all values are reasonably close to a two-choice chance value of 50, and that progressively declining gradients are not initially present.

A second interpretation of the gradients might be that they are the result

of *Ss*' sequential guessing habits. However, since different random orders were used in each block of 35 trials, and the order of presentation of the six blocks was independently randomized for each *S*, it seems unlikely that guessing-habit sequences could exert a systematic bias on the results.

Although the present experiments were conducted in a windowless, black-walled room, illuminated by a single overhead 7.5-w. bulb, all signal lamps, even when unlit, could be rather clearly seen. Because of this, it scarcely seems that the generalization-like gradients could be due to an "inability to discriminate" one lamp from another. Nevertheless, this possibility has been examined by running three additional groups of *Ss* ($N = 26$ per group) under the three levels of room illumination provided by a 7.5-w. lamp, a 60-w. lamp, and a group of lamps totaling 480 w. The data from all three groups yielded clear-cut gradients much like those of Exp. I and II, but no significant differences attributable to room illumination were obtained.

When plots were made of percentages of win responses for each of the six 35-trial blocks, it was found that the gradients began to appear as early as the second block, but were not fully developed until four or five blocks had been completed. The curves show that at first all win percentages were near 50, and that as training continued, the tendency to guess "win" increased progressively for Lamp 4 and decreased for all other lamps. Thus with the empirical frequencies used in Exp. I and II the generalization effect emerged as a differential resistance to the depressing effects (on "win" responses) of losing, rather than a differential heightening. This is one consequence of our procedure that distinguishes it from the typical conditioned-response generalization study.

Summary

This paper describes what is believed to be an improved method for studying spatial generalization with voluntary reactions. The method is designed to avoid difficulties inherent in experiments where false reactions serve as indicants of generalized tendencies. In the technique described, *Ss* are tested in a simulated horse-race-betting situation where seven horses are symbolized by seven lamps spaced at equal intervals along a horizontal line in front of *S*. On each of numerous imaginary races (represented by the lighting of a lamp) *S* must guess whether the horse will "win" or "lose." After he has "placed his bet," *S* is informed by visual signals of the success or failure of the horse. In the experiments described, the middle "horse" actually "won" 80% of its races and all others "won" only 20% of theirs.

Fairly symmetrical, bidirectional, spatial-generalization effects were found in all studies. The lamps nearest the 80% winner at the center evoked significantly more "win" responses than more remotely located lamps, even though objective win frequencies for all but the central lamp were identical. Neither initial preferences nor sequential guessing habits seemed to provide tenable explanations of the results. Moreover, the gradients were unaffected by changes in the discriminability of the stimulus lamps due to wide variations in room illumination.

References

1. Anrep, G. V. The irradiation of conditioned reflexes. *Proc. Roy. Soc.*, 1923, **94,** Series B, 404–425.
2. Bass, M. J., & Hull, C. L. The irradiation of a tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, **17,** 47–65.
3. Bilodeau, E. A., Brown, J. S., & Meryman, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51,** 293–298.
4. Brown, J. S., Bilodeau, E. A., & Baron, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41,** 52–61.
5. Gibson, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24,** 237–253.
6. Mednick, S. A. Distortions in the gradient of stimulus generalization related to cortical brain damage and schizophrenia. *J. abnorm. soc. Psychol.*, 1955, **51,** 536–542.

(Received April 29, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

THE EFFECTS OF DELAY OF REINFORCEMENT UPON AN OPERANT DISCRIMINATION IN THE PIGEON [1]

JEROME L. MYERS[2]

University of Wisconsin

A recent series of experiments by Ferster (1) dealt with the effects of delay of reinforcement upon the rate of the pecking response in the pigeon. He concluded that operant behavior will be sustained, despite delays of reinforcement, as long as 60 sec. if, and only if, the bird develops some consistent mode of responding during the delay interval.

The two experiments reported in this paper deal with an extension of the problem devised by Ferster. In his studies the same color was always present at the start of each trial. The *S* responded to the color by pecking at a transluscent key. The color was then withdrawn and, after a selected interval of delay, *S* was rewarded. In the present group of studies, either a red or yellow light (the samples) illuminated the key at the beginning of each trial and rate of pecking was measured. The delay period followed and was terminated by a reward only if it had been preceded by that color which had been a priori designated as the positive sample. In this situation the possibility arose that the behavior exhibited during the delay interval would be a function of whether the positive or negative sample preceded the delay. There were thus two discriminations which might be learned: red versus yellow and delay-following-red versus delay-following-yellow. The development of these discriminations and the influence of several conditions upon them were investigated.

Experiment I

The purpose of this experiment was to study the development of the discrimination of sample, and to determine whether any changes in behavior during the delay interval could be related to the acquisition of that discrimination. A second objective was to determine whether manipulation of delay interval stimuli affected delay interval behavior, or the discrimination of sample, or both.

Method

Apparatus.—The experimental cage was a box, 10 × 10 × 10 in. The roof was clear Lucite, permitting illumination of the box by overhead fluorescent room lights, and also permitting *E* to observe *S*. The front panel of the box contained a food magazine and an illuminated, 1¼-in. circular hole, behind which was a hinged rectangle of Lucite, the key. When *S* pecked at the key, contact was made with a microswitch lever, thus closing a circuit. On positive trials, pecking at the key opened the food magazine for 4.25 sec.

The recorders, counters, tape transmitter, and relays were all in a room separate from the experimental box. A thick concrete wall prevented the pigeon from obtaining any auditory cues other than those provided by the food magazine.

Procedure.—During the period required to train *S*s in the key-pecking response, the key was illuminated by a white light. On the day before the first experimental session, the key was alternately illuminated by yellow, red, and blue lights. The *S*s received ten 10-sec. fixed-interval reinforcements (i.e., the first peck after 10 sec. was rewarded regardless of the

[1] This article is based on a thesis presented in partial fulfillment of the requirements for the Ph.D. degree, Department of Psychology, University of Wisconsin, 1957. The author wishes to thank F. A. Mote, D. A. Grant, and L. B. Wyckoff, for their helpful advice.

[2] Now at the University of Massachusetts.

pecking rate during the 10 sec.) under each color condition. Ten-second periods of darkness (nonillumination of the key) were interspersed between color trials. Responses during these periods were not reinforced.

At the beginning of any experimental trial the key was illuminated by either a red or a yellow light. Following one of the colors (the positive sample), the first peck to blue at the end of a 10-sec. interval (delay interval) was always reinforced. This response also turned off the blue light illuminating the key. If the delay interval followed the other color stimulus, the key was darkened at the end of 10 sec. and no reinforcement was given. In either event, the key was darkened for 10 sec. between trials, a trial being defined as 10 sec. of red or yellow followed by 10 sec. of blue. Yellow was positive for half the birds, red for the other half.

There were 36 positive and 36 negative trials each day. These were in a prearranged sequence, so that there were never more than three positive or three negative trials in a row and so that positive trials were followed equally often by both positive and negative trials, as were negative trials. This sequence was programmed and presented automatically by means of a Western Union tape transmitter.

The total number of pecks to red, to yellow, to blue following red, and to blue following yellow were recorded on four counters. A Gerbrands cumulative recorder gave a trial-by-trial record of the number of pecks to each stimulus for each *S*.

At the end of each day's session, the number of pecks during each 10-sec. interval was read off the Gerbrands record. Grant's run test (**2**) was applied to the data to determine whether a pigeon was consistently pecking more to the positive than to the negative stimulus and also whether *S* was consistently pecking more to blue following the positive than to blue following the negative stimulus. If the run criterion for sample scores was significant at the .01 level, the experiment was terminated regardless of the results obtained during the delay.

Eight pigeons were run on the problem until the run criterion was met. They were then subdivided into two groups of four. In Group I, after the discrimination on the samples was achieved, the problem was changed so that the key was dark rather than blue during the delay period. This condition was run for only one day since it was felt that if any changes in behavior occurred they would be manifest in the first few trials following the change in delay conditions. In Group II, the four *S*s served as a control group. They were run for three days past the point of discrimination of the sample. Their performance on the first day after the discrimination was achieved could be compared with that of the birds in Group I. This group also presented an opportunity to observe changes in behavior which occurred after the statistical criterion had been reached.

Results

All eight *S*s were able to learn the discrimination of sample. Figure 1 presents the daily pecking totals for the negative sample, the positive sample, delay following the positive sample, and delay following the negative sample, for Group I. The data for Group II present a very similar picture until the point at which a discrimination of sample was achieved. In Fig. 1, this point is indicated by the arrows on the abscissa.

On Day 1′ for Group I a darkened key was substituted for the blue illuminated key during the delay. The rate of pecking during the delay interval dropped off considerably. All discriminations of sample were still present although the run criterion for bird P-1 on Day 1′ was now significant at only the .05 level. No discriminations during the delay were present (i.e., statistically significant) at the time the experiment was terminated. In fact three of the four *S*s in Group I were pecking more to the blue stimulus following the negative sample than to blue following the positive sample on the day discrimination on sample was achieved.

Differences in performances of Groups I and II on Day 1′ were tested. A difference score was obtained for each *S* by subtracting Day 1′ pecks to the negative sample from Day 1′ pecks to the positive sample. These eight scores were then placed in rank order and the Mann-Whitney *U* test (**6**) was employed. There was no significant difference between the

two groups in their discrimination on Day 1′.

The rate of pecking to blue during the delay interval was from three to six times higher than the rate of pecking to either sample. This is predictable since responses during delay were immediately reinforced 50% of the time, and the pigeon was always in position to strike the key when the blue light came on.

To summarize, all *S*s were able to learn the discrimination of sample. No discrimination of delay was manifest at the time the sample discrimination was achieved. Removing the blue light during the delay interval did alter *S*s' behavior during this

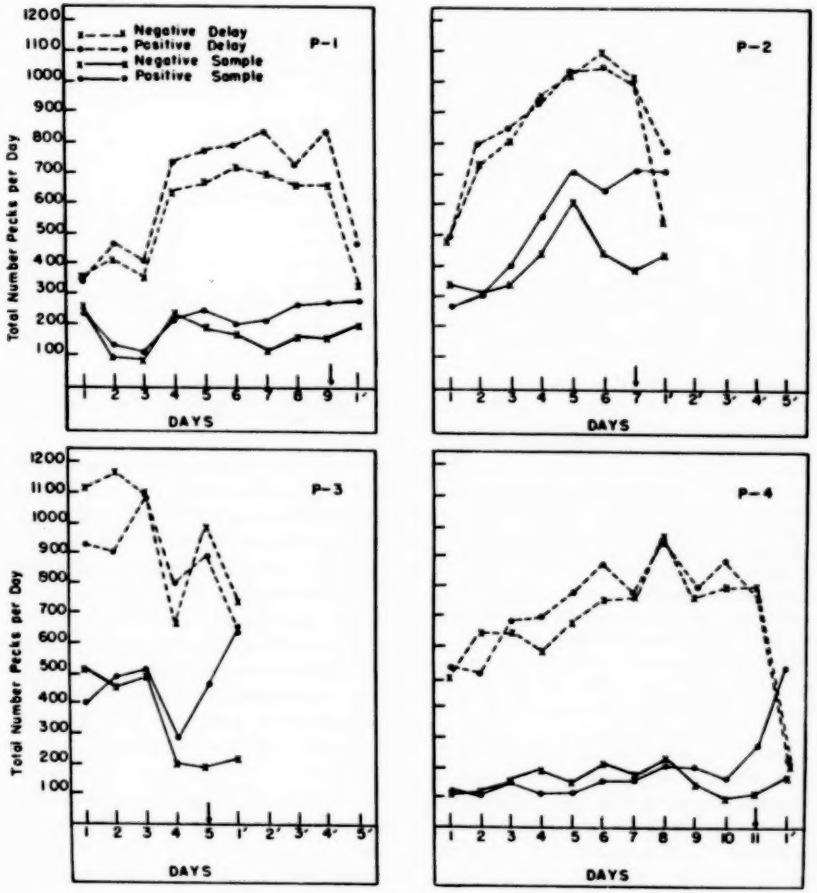


FIG. 1. The number of pecks per day to each stimulus for each *S* in Group I, Exp. I. The arrow on each abscissa indicates the day on which discrimination to sample was first achieved. The primed (′) numbers indicate that the key was darkened during the delay intervals on that day.

interval, but the discrimination of sample did not appear to be affected.

Experiment II

The purpose of Exp. II was to investigate the effects of repeated reversals. Each time *S* learned the discrimination, the situation was changed so that the previously negative stimulus was now positive, and vice versa. One question is whether or not a reversal learning set (**5**) can be established under delay of reinforcement conditions. A second question is whether there is any "learning to learn" phenomenon during the delay period as well. Finally, one may ask whether changes in delay conditions affect the reversal set (if established).

Method

The training and experimental sessions were the same as in Exp. I. Whenever *S* reached the discrimination criterion, reinforcement conditions were reversed so that the previously negative sample was now positive, and the previously positive sample was now negative. These reversals continued until *S* reached the criterion of significance ($P < .05$) on three consecutive days. At this point delay conditions were changed. Reinforcement was no longer contingent upon the occurrence of any given response during the delay, but was immediately presented at the end of the 10-sec. interval following the positive sample. The key was now dark during delay periods. The *S*s were run through three reversal problems under these new conditions. Six naive pigeons were *S*s in this experiment.

Results

Table 1 presents the number of days each *S* required to achieve a discrimination of the red and yellow samples during each successive reversal problem. The primed (′) numbers indicate the three problems run after the discrimination criterion was reached, with the key darkened during delay. Although all *S*s reached the rapid reversal criterion, no consistent trend appears in Table 1. There is some indication that darkening the key did interfere with the ability to rapidly reverse. Substantiation of this must await the running of a control group past the point of three consecutive one-day reversals. Delay conditions were not altered following three successive one-day reversals for Bird P-4 because of apparatus problems which occurred at that point.

All six *S*s developed a discrimination during delay. As shown in Table 1, the first significant delay discrimination usually occurred during the problem just preceding the attainment of the rapid reversal criterion. Figure 2 is based on the

TABLE 1

Number of Days Required by Each Pigeon to Reach Criterion on Each Reversal Problem

Reversal Problem	Pigeons P-1	P-2	P-3	P-4	P-5	P-6
0	9	12	6	4	6	8
1	5	10	5	14	5	16
2	8	8	10	6*	5	8
3	4	7	10	1	5	7
4	4	5*	4	1	7	7
5	4	1	5	1	3	6
6	7	1	2*		4	7
7	4	1	2		4	4
8	6		2		6	4*
9	1		2		2	1
10	2*		1		5	1
11	1		4		4	1
12	1		1		5	
13	1		1		1	
14			1		2	
15					1	
16					3	
17					2	
18					4*	
19					1	
20					1	
21					1	
1′	1	1	3		5	2
2′	1	2	1		1	1
3′	1	1	1		1	2

* A significant discrimination during delay first appeared during the running of this problem.

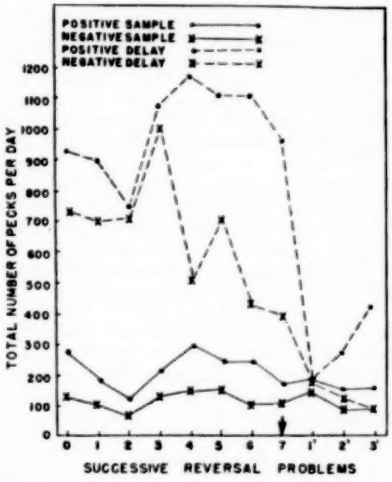


FIG. 2. The total number of pecks for one *S* to each stimulus on the day the reversal criterion was met. The primed (') numbers indicate that the key was darkened during these reversal problems. The arrow on the abscissa shows the point at which the rapid reversal criterion was met.

record of one *S* and shows the total number of pecks under each condition for the day each discrimination problem was learned. The delay discrimination is apparently the result of a sharp drop in responses to blue following the negative sample. Once learned, the discrimination of blue is reversed as readily as the discrimination of sample.

To summarize, a reversal learning set, with regard to the samples, was established. Discrimination of the delay-following-red versus delay-following-yellow appeared at about the same time as did rapid reversals on the sample. This differential rate of pecking was also rapidly reversed as a function of which color sample led to reinforcement. The reversal set may have been somewhat affected by changes in the delay-interval stimulus color.

DISCUSSION

Despite delays of reinforcement of 10-sec. duration, all pigeons were able to learn the discrimination. This is of interest in view of the fact that no differential cues were consistently paired with the reward. The box, the color of the key, and the position of the key were all the same following both color samples. Gulde (**4**) and Riesen (**7**), controlling differential cues during delay, found that rats and chimpanzees could not solve problems involving delays of 5 sec. The difference between these two studies and the present one would seem to lie in the complexity of the learning task, rather than in the factor of delay stimulation. More than one possible discriminable dimension was present in the Gulde and Riesen studies, whereas the pigeons were concerned only with the dimension of color.

When the pigeon's problem became the more complicated one of forming a reversal learning set, the solution of the problem was preceded by the formation of a discrimination during the delay. It is possible that, as a problem becomes more complex, there is a greater need for differential cues, and that a high rate of delay-interval pecking, associated with the reinforcement, serves this purpose.

The present hypothesis is that *all* delayed reward learning does not require differential stimulation during delay, and that the relationship between learning and the need for differential cues is largely a function of the complexity of the learning task. This may be contrasted with the viewpoint put forth by Spence in 1947 (**8**). He hypothesized that immediate reinforcement must be present in order for learning to occur, that if this reinforcement is not of a primary nature, it must be secondary reinforcement, and therefore discrimination learning under delay conditions necessitates differential external or internal cues.

It is possible that some differential behavior, more difficult to observe than rates of pecking, was present during

the delay intervals. Perhaps some unobserved or unobservable response, such as salivation following the positive sample, may have been present and served to mediate the discrimination. A response of this sort seems rather unsatisfactory as an explanation and leaves one still wondering why the obvious secondary reinforcer, proprioceptive stimulation from pecking, was not differentiated.

Although differential delay stimulation may not be essential for discrimination learning, Grice (3) has shown that it does facilitate learning. This is also suggested by the results of Exp. II in this paper. Another question that merits attention is whether or not a level of performance, reached with the help of cues during delay, can be sustained if the cues are removed. Other variables which may prove important in delayed-reward learning are the particular stimuli present during the delay and the previous learning history of *S*.

Summary

Two experiments on delayed-reward learning in the pigeon were performed in a Skinner Box. On each trial either a red or yellow light illuminated a key for 10 sec. and *S*'s rate of pecking at the key was measured. A 10-sec. period during which a blue light illuminated the key always followed, regardless of which color had been originally presented. This delay period was terminated by a food reward only if it had been preceded by red (or yellow) which had been a priori designated as the positive stimulus. Seventy-two trials were run each day until *S* reached a criterion for discrimination of red and yellow.

In Exp. I, four *S*s were run until the discrimination criterion was reached, and then the key was not illuminated during the delay interval. Four other *S*s were run for three days without change after the discrimination criterion was reached (Control Group) The change in the delay condition did not appear to have any effect upon the attained discrimination.

In Exp. II, six *S*s were tested for reversal learning. Each time the statistical criterion was reached, reinforcement was shifted to the alternate color sample. This procedure was continued until the discrimination was attained on three successive days. Three additional reversals were run with the key dark during the delay interval. A sudden decrease in the number of trials needed to reach criterion occurred at about the same time that *S*s showed differential pecking behavior during the delay interval. This differential behavior dropped out when the delay condition was altered, and the reversal sets appeared to be impaired.

The role of differential cues in delayed-reward learning was discussed, and it was suggested that such cues are not always necessary for delayed-reward learning. The complexity of the learning task was suggested as an important variable which requires investigation.

References

1. Ferster, C. B. Sustained behavior under delayed reinforcement. *J. exp. Psychol.*, 1953, **45**, 218–224.
2. Grant, D. A. Additional tables of the probability of "runs" of correct responses in learning and problem solving. *Psychol. Bull.*, 1947, **44**, 276–279.
3. Grice, G. R. The relation of secondary reinforcement to delayed reward in visual discrimination learning. *J. exp. Psychol.*, 1948, **38**, 1–16.
4. Gulde, C. J. The effects of delayed reward on the learning of a black-white discrimination problem by the albino rat. Unpublished Master's thesis, Univer. of Iowa, 1941.
5. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
6. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
7. Riesen, A. H. Delayed reward in discrimination learning by chimpanzees. *Comp. Psychol. Monogr.*, 1940, **15**, 1–53.
8. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.

(Received April 29, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

AN ANALYSIS OF TWO-PERSON GAME SITUATIONS IN TERMS OF STATISTICAL LEARNING THEORY[1]

RICHARD C. ATKINSON[2] AND PATRICK SUPPES

Applied Mathematics and Statistics Laboratory, Stanford University

This study represents an extension of statistical learning theory to a class of two-person, zero-sum game situations. Because the theory has been mainly developed in connection with experiments dealing with individual learning problems, its predictive success in an experimental area involving interaction between individuals provides an additional measure of the scope of its validity. It should be emphasized that the study reported here was not conceived as providing an empirical test of the adequacy of learning theory as opposed to game theory; although we use the language of game theory to describe the study, the game characteristics of the situation were not apparent to *S*s. This point is amplified below.

For the purposes of this experiment, a play of a game is a trial. On a given trial each of the two players independently makes a choice between one of two alternatives—that is, he makes one of two possible responses. After the players have indicated their choices, the outcome of the trial is announced to each player.

On all trials, the game is described by the following pay-off matrix.

	B_1	B_2
A_1	a_1	a_2
A_2	a_3	a_4

The players are designated A and B. The responses available to A are A_1 and A_2; similarly, the responses available to B are B_1 and B_2. If A selects A_1 and B selects B_1 then there is a probability a_1 that A is "correct" and B is "incorrect," and a probability $1 - a_1$ that A is "incorrect" and B is "correct." These two joint events are exhaustive since it is required that exactly one player is correct on each trial. The outcomes of the other three response pairs are identically specified in terms of a_2, a_3 and a_4.

The interaction of the players is limited by two factors: (*a*) neither player is shown the pay-off matrix, (*b*) neither player is directly informed of the responses of the other player. Thus, from the standpoint of the general theory of rational behavior (**4**), *S* should not regard himself as playing a 2×2 game with known pay-off matrix but should view the situation as a multi-stage decision problem against an unknown opponent. However, selection of an optimal strategy in this multi-stage decision problem is far from a trivial task mathematically, and it is scarcely to be expected that any *S* would use such a strategy. The virtue of statistical learning theory is that it yields a quantitative prediction of how organisms actually do behave in such situations.

Our theoretical analysis of the behavior of *S*s in the situation described is based on two distinct but closely related models. Since a detailed mathematical analysis of these models will be presented elsewhere, the present statement will concern only the most salient facts and omit mathematical proofs.

Linear model.—The first model is an extension of a linear model developed by Estes and Burke (**6**). Experimental tests of this formulation for one-person learning situations have been reported

[1] This research was supported by the Behavioral Sciences Division of the Ford Foundation and by the Group Psychology Branch of the Office of Naval Research. The authors are indebted to W. K. Estes for several stimulating discussions of the ideas on which this experiment is based.

[2] Now at the University of California, Los Angeles.

(2, 9, 13). The basic assumption of the model is that response probability on a given trial is a linear function of the probability on the preceding trial. When a response is reinforced its probability increases; the reinforcement of any other response decreases its probability.

In the present situation, where two responses are available to each S, if a response occurs and is designated as "correct," then the response is reinforced; if a response occurs and is designated as "incorrect," then the alternative response is reinforced. More specifically, let α_n be the probability of response A_1 on Trial n. The rules of change are: (a) if A_1 is reinforced on Trial n, then $\alpha_{n+1} = (1 - \theta_A)\alpha_n + \theta_A$; ($b$) if A_2 is reinforced on Trial n then $\alpha_{n+1} = (1 - \theta_A)\alpha_n$, where $0 < \theta_A \leq 1$. Identical rules are specified for β_n, the probability of a B_1 response, in terms of θ_B.

The following pair of recursive equations can then be derived for the mean probabilities $\bar{\alpha}_n$ and $\bar{\beta}_n$, where $\bar{\gamma}_n$ is the mean probability of the joint event that on Trial n Player A will make response A_1 and Player B response B_1.

$$\bar{\alpha}_{n+1} = (1 - 2\theta_A + \theta_A a_2 + \theta_A a_4)\bar{\alpha}_n + \theta_A(a_4 - a_3)\bar{\beta}_n + \theta_A(1 - a_4) + \theta_A(a_1 + a_3 - a_2 - a_4)\bar{\gamma}_n$$

$$\bar{\beta}_{n+1} = (1 - \theta_B a_3 - \theta_B a_4)\bar{\beta}_n + \theta_B(a_2 - a_4)\bar{\alpha}_n + \theta_B a_4 + \theta_B(a_3 + a_4 - a_1 - a_2)\bar{\gamma}_n.$$

It may be shown that $\bar{\alpha}$, $\bar{\beta}$ and $\bar{\gamma}$, the asymptotic probabilities in the sense of Cesaro **(11)**,[3] exist and are independent of the initial probabilities α_0, β_0, γ_0. However, in general these asymptotic quantities depend on θ_A and θ_B, and no simple results are obtainable for the quantities individually. On the other hand, an interesting linear relation between $\bar{\alpha}$ and $\bar{\beta}$, which is independent of $\bar{\gamma}$, θ_A and θ_B, can be derived, namely:

$$[(a_3 + a_4 - a_1 - a_2) + (a_1 a_2 - a_3 a_4)]\bar{\alpha} = (a_1 a_3 - a_2 a_4)\bar{\beta} + \tfrac{1}{2}(a_3 + a_4 - a_1 - a_2) + a_4(a_2 - a_3). \quad (1)$$

The line determined by this equation has been labeled the *interaction line* since the exact point on the line specifying the asymptotic probabilities $\bar{\alpha}$ and $\bar{\beta}$ is a function of both θ_A and θ_B. It is interesting to observe that in the corresponding one-person learning situation, the interaction line degenerates to a point, while in the three-person situation an interaction surface is obtained.

Finite Markov model.—In this model the simplifying assumption is made that on all trials a player's response behavior is determined by a single stimulus—that is, the event associated with the onset of a trial. The S is described as being in one of two possible states: (a) if in State 1, the stimulus is conditioned to Response 1 and, in the presence of the stimulus, Response 1 will be elicited; (b) if in State 2, the stimulus is conditioned to Response 2 and, in the presence of the stimulus, Response 2 will be elicited. Thus, on any Trial n, the two players are described in terms of one of the following four states: <1,1>, <1,2>, <2,1> and <2,2> where the first member of a couple indicates the state of Player A and the second, the state of Player B. For example, <2,1> means that Player A will make response A_2 and Player B will make response B_1. It is postulated that the change of states from one trial to the next is Markovian, and the following analysis is used to derive the transition matrix **(10, 11)** of the process.

When one of Player A's responses is reinforced on Trial n there is (a) a probability θ_A that the stimulus governing Player A's response will be conditioned to the reinforced response and therefore, on Trial $n + 1$ Player A will make the response reinforced on Trial n and, (b) a probability $1 - \theta_A$ that the conditioned status of the stimulus will remain unchanged and therefore, on Trial $n + 1$ Player A will repeat the response made on Trial n. Identical rules describe the process for Player B in terms of θ_B.[4]

[3] To be explicit,

$$\bar{\alpha} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \bar{\alpha}_i,$$

and similarly, for $\bar{\beta}$ and $\bar{\gamma}$.

[4] The Markov process derived from these assumptions differs in certain respects from that which can be derived from the Estes and Burke stimulus sampling model **(6)**. In their formulation the stimulus is conceptualized as being

For this set of assumptions and the pay-off probabilities a_1, a_2, a_3 and a_4, the transition matrix describing the learning process can be derived and is as follows:

	<1,1>	<1,2>	<2,1>	<2,2>
<1,1>	$a_1(\theta_A - \theta_B) + (1 - \theta_A)$	$a_1\theta_B$	$(1 - a_1)\theta_A$	0
<1,2>	$a_2\theta_B$	$a_2(\theta_A - \theta_B) + (1 - \theta_A)$	0	$(1 - a_2)\theta_A$
<2,1>	$(1 - a_3)\theta_A$	0	$a_3(\theta_A - \theta_B) + (1 - \theta_A)$	$a_3\theta_B$
<2,2>	0	$(1 - a_4)\theta_A$	$a_4\theta_B$	$a_4(\theta_A - \theta_B) + (1 - \theta_A)$

Rows designate the state on Trial n and columns the state on Trial $n + 1$. Thus $(1 - a_3)\theta_A$, the entry in Row 3, Column 1, is the conditional probability of being in State <1,1> on Trial $n + 1$ given that the pair of *S*s was in State <2,1> on Trial n, because:

$$(1-a_3)\theta_A=\theta_A\theta_B(1-a_3)+\theta_A(1-\theta_B)(1-a_3)+(1-\theta_A)\theta_B\cdot 0+(1-\theta_A)(1-\theta_B)\cdot 0.$$

From these one-stage transition probabilities an explicit solution is obtained for the Cesaro asymptotic probabilities of an A_1 and B_1 response; as in the case of the linear model these quantities are denoted as $\bar{\alpha}$ and $\bar{\beta}$, respectively. The general equations for $\bar{\alpha}$ and $\bar{\beta}$ are too lengthy to reproduce here but certain results are noteworthy. It can be shown that $\bar{\alpha}$ and $\bar{\beta}$ are related by the identical interaction line determined by Equation 1 of the linear model. For the Markov model, however, it can in addition be proved that the point on the interaction line describing a particular pair of *S*s' asymptotic behaviors is uniquely determined by the ratio of θ_A to θ_B. Further, even without a knowledge of the specific values of θ_A and θ_B one can specify a fairly narrow interval on the interaction line within which $\bar{\alpha}$ and $\bar{\beta}$ must fall by taking the limits of $\bar{\alpha}$ and $\bar{\beta}$ as the ratio θ_A/θ_B approaches zero or becomes large.

Particular cases of the theoretical analysis may be illustrated by examining predictions for the parameter values employed in this experiment. Three sets of a_i values were used corresponding to three classical cases of 2×2 games in the theory of zero-sum, two-person games (**12**).

The first case is labeled the *Mixed Group*, since both players have mixed minimax strategies. The a_i values are given by the pay-off matrix

	B_1	B_2
A_1	$\frac{3}{4}$	1
A_2	$\frac{1}{2}$	$\frac{1}{4}$

The minimax strategy for Player A is to choose A_1 with probability $\frac{3}{5}$, and the minimax strategy for B is to choose B_1 with probability $\frac{5}{6}$. In the Markov model

$$\bar{\alpha} = .600 \tag{2}$$

$$\bar{\beta} = \frac{35(\theta_A/\theta_B) + 22}{50(\theta_A/\theta_B) + 40}. \tag{3}$$

Note that $\bar{\alpha}$ is independent of θ_A/θ_B. From Equation 3 one obtains as bounds on $\bar{\beta}$:

$$.550 < \bar{\beta} < .700. \tag{4}$$

If one assumes $\theta_A = \theta_B$, then $\bar{\beta} = .633$. For this case the interaction line is the line satisfying Equation 2.

The second case is labeled the *Pure Group*, since both players have pure minimax strategies. The particular values

composed of a *large* number of stimulus elements each of which is sampled with probability θ and, once sampled, conditioned to the reinforced response with probability 1. Further, the probability of a response is defined as the proportion of stimulus elements in the sample conditioned to the response. In the model used in this paper it is assumed that the single stimulus is sampled on each trial with probability 1.

are given by the matrix

	B_1	B_2
A_1	$\frac{1}{2}$	1
A_2	$\frac{1}{2}$	$\frac{1}{4}$

Here $a_1 = \frac{1}{2}$ is a saddle point of the matrix and from the standpoint of game theory the optimal strategy for Player A is to play A_1 with probability 1 and for B to play B_1 with probability 1. In the Markov model

$$= .667 \tag{5}$$

$$\bar{\beta} = \frac{6(\theta_A/\theta_B) + 5}{9(\theta_A/\theta_B) + 9}. \tag{6}$$

As in the previous case, $\bar{\alpha}$ is independent of θ_A/θ_B and the interaction line is the line satisfying Equation 5. From Equation 6 one obtains as bounds on $\bar{\beta}$:

$$.555 < \bar{\beta} < .667. \tag{7}$$

If one assumes that $\theta_A = \theta_B$, then $\bar{\beta} = .611$.

The third case is labeled the *Sure Group* since both players have sure-thing strategies (i.e., given the pay-off matrix one of the two responses available to each player is at least as good or better than the other response regardless of what his opponent does). The parameter values are given by the matrix

	B_1	B_2
A_1	$\frac{1}{2}$	1
A_2	$\frac{1}{4}$	$\frac{3}{4}$

The sure-thing strategies for Players A and B are A_1 and B_1 respectively. In the Markov model

$$\bar{\alpha} = \frac{5(\theta_A/\theta_B) + 15}{7(\theta_A/\theta_B) + 23} \tag{8}$$

$$\bar{\beta} = \frac{5(\theta_A/\theta_B) + 16}{7(\theta_A/\theta_B) + 23}, \tag{9}$$

and as bounds one has:

$$.652 < \bar{\alpha} < .711 \tag{10}$$

$$.696 < \bar{\beta} < .711. \tag{11}$$

If one assumes that $\theta_A = \theta_B$, then $\bar{\alpha} = .667$ and $\bar{\beta} = .700$. For this case the interaction line is determined by the equation:

$$3\bar{\alpha} = 10\bar{\beta} - 5. \tag{12}$$

Method

Subjects.—The *Ss* were 120 undergraduates obtained from introductory courses in psychology and philosophy at Stanford University. They were randomly assigned to the Mixed, Pure, and Sure Groups with the restriction that there were 20 pairs of *Ss* in each group.

Apparatus.—The *Ss*, run in pairs, sat at opposite ends of an 8 × 3-ft. table. Mounted vertically on the table top facing each *S* was a 50-in. wide by 30-in. high black panel placed 22 in. from the end of the table. The *E* sat between the two panels and was not visible to either *S*. The apparatus, as viewed from *S*'s side, consisted of two silent operating keys mounted 8 in. apart on the table top and 12 in. from the end of the table; upon the panel, three milk-glass panel lights were mounted. One of these lights, which served as the signal for *S* to respond, was centered between the keys at a height of 17 in. from the table top. Each of the two remaining lights, the reinforcing signals, was at a height of 11 in. directly above one of the keys. The presentation and duration of the lights were automatically controlled. The *Ss* were not visible to one another and could not see each other's responses or panel lights.

Procedure.—The *Ss* were read the following instructions: "We always run *Ss* in pairs because this is the way the equipment has been designed and also because it is the most economical procedure. Actually, however, you are both working on two completely different and independent problems.

"The experiment for each of you consists of a series of trials. The top center lamp on your panel will light for about 2 sec. to indicate the start of each trial. Shortly thereafter one or the other of the two lower lamps will light up. Your job is to predict on each trial which one of the two lower lamps will light and indicate your prediction by pressing the proper key. That is, if you expect the left lamp to light press the left key, if you expect the right lamp to light press the right key. On each trial press one or the other of the two keys but never both. If you are not sure which key to press then guess.

"Be sure to indicate your choice by pressing the proper key immediately after the onset of the signal light. That is, when the signal light goes on press one or the other key down and release it. Then wait until one of the lower lights goes on. If the light above the key you pressed goes on your prediction was correct, if the light above the key opposite from the one you pressed goes on you were incorrect, and should have pressed the other key. At times

you may feel frustrated or irritated if you cannot understand what the experiment is all about. Nevertheless, continue trying to make the very best prediction you can on each trial."

For each pair of *S*s, one was randomly selected as Player A and the other as Player B. Further, for each *S* one of the two response keys was randomly designated Response 1 and the other Response 2 with the restriction that the following possible combinations occurred equally often in each of the three experimental groups: (*a*) A_1 and B_1 on the right, (*b*) A_1 on the right and B_1 on the left, (*c*) A_1 on the left and B_1 on the right, and (*d*) A_1 and B_1 on the left.

Following the instructions, 200 trials were run in continuous sequence. For each pair of *S*s sequences of reinforcing lights were generated in accordance with assigned values of a_1 and observed responses.

On all trials the signal light was lighted for 3.5 sec.; the time between successive signal exposures was 10 sec. The reinforcing light followed the cessation of the signal light by 1.5 sec. and remained on for 2 sec.

At the end of the session each *S* was asked to describe what he thought was involved in the experiment. Only one *S* indicated that he believed the reinforcing events depended in any way on a relationship between his responses and the other player's responses. His record and that of his partner were eliminated from the analysis and replaced by another pair.

Results and Discussion

Mean learning curves and asymptotic results.—Figure 1 provides a description of behavior over all trials of the experiment. In this figure the mean proportions of A_1 and B_1 responses in successive blocks of 40 trials are given for the sequence of 200 trials. An inspection of this figure indicates that responses are fairly stable over the last 100 trials except possibly for B_1 responses in the Pure Group. To check the stability of response behavior for individual data, *t*'s for paired measures were computed between response proportions for the first and last halves of the final block of 60 trials. In all cases the obtained values of *t* fall short of significance at the .05 level.

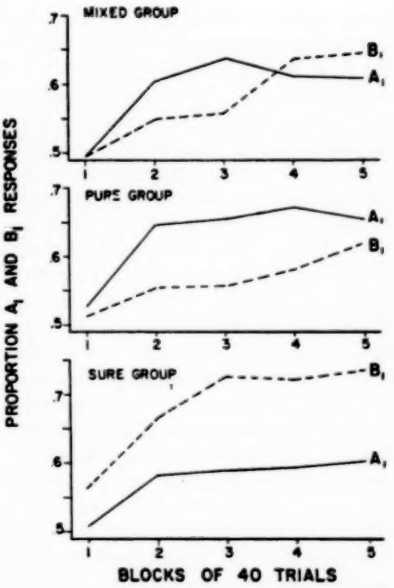


Fig. 1. Observed proportions of A_1 and B_1 responses in blocks of 40 trials for the three experimental groups.

It appears reasonable to assume that a constant level of responding has been reached; consequently the proportions computed over the last 60 trials were used as an estimate of $\bar{\alpha}$ and $\bar{\beta}$. Table 1 presents the observed mean proportions of A_1 and B_1 responses in the last 60 trial block and the *SD*'s associated with these means.

TABLE 1

Predicted and Observed Mean Proportions of A_1 and B_1 Responses Over the Last Block of 60 Trials

	Response A_1			Response B_1		
	Pred.	Obs.	*SD*	Pred.	Obs.	*SD*
Mixed	.600	.605	.0794	.633	.649	.0874
Pure	.667	.670	.0832	.611	.602	.0634
Sure	.667	.606	.1005	.700	.731	.0760

Each entry is based on $N = 20$. The values predicted by the Markov model for $\theta_A = \theta_B$ are also presented.

Inspection of Table 1 indicates that predicted and observed results are extremely close for the Mixed and Pure Groups; t tests of the difference between these values do not approach significance at the .05 level. For the Sure Group the difference between Player B's observed and theoretical values is also not significant; but for Player A, the difference is significant. Specifically, the observed proportion of A_1 responses for the Sure Group is less than predicted. Note, however, that one may relinquish the assumption that $\theta_A = \theta_B$ and, given the boundary conditions specified by Equations 10 and 11, determine for the Sure Group the point on the interaction line (see Equation 12) which is nearest the observed point. This nearest point is $\bar{\alpha} = .652$ and $\bar{\beta} = .696$. For this point the difference between observed and theoretical values is not significant at the .05 level for either A_1 or B_1 responses.

Game theory comparisons.—It is of interest to compare observed values with the game-theoretic optimal strategies discussed earlier, for it can be reasonably maintained that even though Ss do not know the pay-off matrix, after a large number of trials they have learned enough about the situation to approach an optimal game strategy. Concerning such a conjecture, the results for the Pure and Sure Groups seem decisive: the optimal game strategies of responding A_1 or B_1 with probability 1, for Player A or B, respectively, are not even crudely approximated by the observed means. Moreover, the maximum individual value in each group of 20 Ss does not approach 1; for the Pure Group max α is .80 and max β is .71, while for the Sure Group max α is .77 and max β is .84.

The results for the Mixed Group also fail to support the hypothesis that Ss, in the long run, will approach an optimal game strategy. The observed $\bar{\alpha}$ of .605 and $\bar{\beta}$ of .649 both differ significantly from their respective minimax strategies of $\frac{1}{3}$ and $\frac{2}{3}$ at beyond the .001 level.

Several questions are suggested by these comparisons with game theory that are pertinent to a theory of small groups. First, would the learning theory predictions be less applicable and the optimal game strategies more closely approximated if Ss are explicitly told that they are competing with each other? Subsequent experimental work (**3**) indicates that the answer to this question is probably negative. Second, would optimal game strategies be more closely approximated if Ss were run for a very large number of trials over a period of several days? What evidence there is on this question from individual learning situations (**7, 8, 9**) tends to support the hypothesis that the long run mean probabilities would stay close to the learning theory predictions. However, detailed experimental investigation would be worth while. Third, would the present experimental results be affected if Ss were paid for correct responses and penalized monetarily for incorrect responses? The models formulated in the first part of this paper are not rich enough in conceptual content to express formally possible effects of different types of reinforcing events. Fourth, will the obvious generalization of the two models to the interaction of more than two Ss be experimentally substantiated, and how will observed response probabilities compare with various proposed "solutions" of n-person games?

Adequacy of Markov model.—Because of the relatively simple mathematical character of stationary Markov processes with a finite number of

TABLE 2

OBSERVED TRANSITION MATRICES CORRESPONDING TO THE THEORETICAL TRANSITION MATRIX SPECIFIED BY THE MARKOV MODEL

(Computed over the last 100 Trials)

	Mixed Group				Pure Group				Sure Group			
	<1,1>	<1,2>	<2,1>	<2,2>	<1,1>	<1,2>	<2,1>	<2,2>	<1,1>	<1,2>	<2,1>	<2,2>
<1, 1>	.37	.22	.30	.11	.38	.27	.24	.11	.43	.18	.29	.10
<1, 2>	.54	.25	.15	.06	.50	.31	.11	.08	.52	.19	.22	.07
<2, 1>	.35	.16	.30	.19	.30	.20	.29	.21	.47	.12	.31	.10
<2, 2>	.28	.34	.17	.21	.31	.36	.16	.17	.27	.17	.37	.19

states, it is possible to ask certain detailed questions about our data from the standpoint of Markov models. Probably the most direct question is: How do the aggregate transition matrices for each of the three experimental groups compare with the theoretical transition matrix derived in the first part of the paper? Table 2 presents the observed matrices computed over the last 100 trials. Since each group contained 20 pairs of *S*s, each matrix is based on 2000 observations. No statistical test is needed to see that the observed matrices differ significantly from the theoretical matrix. It is sufficient to observe that in the theoretical matrix (for all sets of parameter values a_i) the antidiagonal is identically zero, but in the observed matrices every entry on the antidiagonals is markedly different from zero. As a matter of fact, it would be surprising to find a very close agreement between the theoretical and observed matrices, for the theoretical matrix was derived from exceedingly simple assumptions. In particular, it seems unlikely that the detailed pattern of *S*s' responses could be accounted for by a model which conceptualizes the effective stimulus as a single element identical from trial to trial.

For experimental situations involving more than one *S* even the extension to a two-element stimulus model is not trivial from the standpoint of computing the simplest quantities desired—namely, asymptotic response probabilities. For example, if in the one-element model one identifies the stimulus element as the signal light, one natural two-element model is to identify two successive signals as the two stimuli. The Markov process derived from this assumption has, for the present experiment, 16 states.

Fortunately, without examining a specific two-stage Markov model one can ask one highly relevant question about the present data: Can the data be more adequately accounted for by a two-stage model which employs information about the *S* on the previous two trials as compared with a one-stage model which employs information about only one preceding trial? For this purpose the χ^2 test described by Anderson and Goodman (**1**) was used. The null hypothesis is that $p_{ijk} = p_{jk}$ for $i = 1, \cdots, 4$ where p_{ijk} is the probability of State k given successively States i and j on the two previous trials and p_{jk} is the probability of State k simply given State j on the preceding trial. To test this hypothesis the following sum was

computed for aggregate group data:

$$\chi^2 = \sum_{i,j,k} n_{ij}^*(\hat{p}_{ijk} - \hat{p}_{jk})^2/\hat{p}_{jk},$$

where $n_{ij}^* = \sum_k n_{ijk}$. If the null hypothesis is true, χ^2 has the usual limiting distribution with $4(4-1)^2 = 36\,df$.

The obtained values of χ^2 were 81.8 for the Mixed Group, 50.5 for the Pure Group and 52.9 for the Sure Group. For the Pure and Sure Groups the value of χ^2 is not significant at the .05 level, for the Mixed Group it is. Independent of any specific model these results indicate that for two of the three groups the learning process is fairly well approximated by a one-stage Markov process. Moreover, it is to be noted that the significant χ^2 for the aggregated data of the Mixed Group does not entail that individual pairs of *S*s were not one-stage Markovian in their responses, for the sum (in the sense pertinent here) of several Markov processes is not necessarily a Markov process. The relatively small number of observations for a given pair of *S*s ruled out a separate χ^2 test for each pair.

If, on the other hand, one accepts the approximate one-stage Markovian character of the learning process studied in this experiment, and asks if this process is stationary in the sense that the observed transition probabilities are constant over all trials, the answer is negative. In a χ^2 analysis (**1**) of the aggregate observed transition matrix for the first 100 trials compared with the last 100 trials, the difference was significant at the .01 level for all three groups. These results suggest that nonstationary, single-element models need to be explored in addition to an analysis of stationary multi-stimulus element models.

Observed and predicted variances in the linear model.—The close agreement between predicted and observed mean asymptotic responses suggests a check of the linear model against another measure of behavior. Specifically, we were interested in checking the variance predicted by the model against the experimental results on variability presented in Table 1. The observed *SD*'s in this table relate to the proportions of A_1 and B_1 responses in blocks of 60 trials; comparable theoretical quantities will be designated $\sigma(A_1)$ and $\sigma(B_1)$, respectively.

Unfortunately direct analytical computation of $\sigma(A_1)$ and $\sigma(B_1)$ seems impossible. Consequently it was necessary to resort to "Monte Carlo methods" (**5**). The basic idea of the approach is to construct a *system* which follows the rules specified by the theory and then make observations on the behavior of the system. By taking a large number of such observations one obtains precise estimates of theoretical quantities. In the present case, what might be considered a hypothetical *S* was built by programming an I.B.M. Type 650 digital computor so that its sequence of commands corresponded to the operations specified by the linear model.

Employing this procedure, estimates of $\sigma(A_1)$ and $\sigma(B_1)$ were obtained for various values of θ_A and θ_B. Results from other experiments (**2, 9, 13, 14**) suggested that θ values for the present study would undoubtedly be bounded between .01 and .50. Hence, combinations of .01, .10, and .50 were used in the computation; a finer gradation of values would have been desirable but the cost of computor time made this prohibitive. The results of the Monte Carlo runs are presented in Table 3.

TABLE 3

MONTE CARLO ESTIMATES OF $\sigma(A_1)$ AND $\sigma(B_1)$ FOR VARIOUS VALUES OF θ_A AND θ_B

θ_A	θ_B	Mixed Group		Pure Group		Sure Group	
		$\sigma(A_1)$	$\sigma(B_1)$	$\sigma(A_1)$	$\sigma(B_1)$	$\sigma(A_1)$	$\sigma(B_1)$
.01	.01	.030	.031	.033	.024	.024	.021
.01	.10	.021	.042	.034	.051	.021	.047
.01	.50	.027	.061	.026	.053	.030	.046
.10	.01	.049	.031	.058	.022	.051	.021
.10	.10	.050	.052	.059	.036	.050	.035
.10	.50	.060	.057	.060	.043	.065	.053
.50	.01	.055	.024	.076	.018	.071	.021
.50	.10	.062	.037	.064	.051	.049	.045
.50	.50	.066	.045	.071	.046	.074	.036

A comparison of Tables 1 and 3 indicates that, for all cases, the observed variability is greater than predicted by the model. Even the most favorable comparisons between observed and predicted values prove to be significantly different at the .05 level when a χ^2 test of variances is employed. The finding that the linear model tends to underestimate observed variability is not surprising in view of similar results from other experiments employing linear operator models to account for individual learning data.

SUMMARY

The study deals with an analysis of a zero-sum, two-person game situation in terms of statistical learning theory and game theory.

The *Ss* were run in pairs for 200 trials. A single play of the game is treated as a trial. On a trial each player makes a choice between one of two alternative responses; after the players have made their response, the outcome of the trial is announced. The responses available to Player A are designated A_1 and A_2; similarly the responses available to Player B are B_1 and B_2. If Player A selects A_1 and Player B selects B_1, then there is a probability a_1 that Player A is "correct" and Player B is "incorrect" and a probability $1 - a_1$ that Player B is "correct" and Player A is "incorrect." The outcome of the other three response pairs is identically specified in terms of a_2, a_3, and a_4. The *Ss* were instructed to maximize the number of correct responses.

Three groups were run, each employing a different set of a_i values. The selection of these values was determined by game-theoretic considerations; that is, a group had available either a sure-thing strategy, a pure minimax strategy, or a mixed minimax strategy.

Analysis of the data was in terms of two different but related stochastic models for learning and game theory. Specifically the following detailed comparisons of data and theory were made: (*a*) mean asymptotic response probabilities, (*b*) one- and two-stage transition probabilities, and (*c*) variances associated with asymptotic response probabilities.

REFERENCES

1. ANDERSON, T. W., & GOODMAN, L. A. Statistical inference about Markov chains. *Ann. math. Statist.*, 1957, **28**, 89–110.
2. ATKINSON, R. C. An analysis of the effect of nonreinforced trials in terms of statistical learning theory. *J. exp. Psychol.*, 1956, **52**, 28–32.
3. ATKINSON, R. C., & SUPPES, P. An analysis of a two-person interaction situation in terms of a Markov process. Tech. Rep. No. 9, Contract NR 171-034, Applied Mathematics & Statistics Laboratory, Stanford Univer., 1957.
4. BLACKWELL, D., & GIRSHICK, M. A. *Theory of games and statistical decisions.* New York: Wiley, 1954.
5. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning.* New York: Wiley, 1955.
6. ESTES, W. K., & BURKE, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

7. Estes, W. K., & Burke, C. J. Application of a statistical model to simple discrimination learning in human subjects. *J. exp. Psychol.*, 1955, **50**, 81–88.
8. Estes, W. K., Burke, C. J., Atkinson, R. C., & Frankmann, J. P. Probabilistic discrimination learning. *J. exp. Psychol.*, 1957, **54**, 233–239.
9. Estes, W. K., & Straughan, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.
10. Feller, W. *An introduction to probability theory and its applications.* New York: Wiley, 1950.
11. Frechet, M. *Recherches théoriques modernes sur le calcul des probabilités*, Vol. 2, Paris: Gauthier-Villars, 1938.
12. McKinsey, J. C. C. *Introduction to the theory of games.* New York: McGraw-Hill, 1952.
13. Neimark, E. D. Effect of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.
14. Toda, M. Guessing sequence under various conditions of payoff. *Jap. psychol. Res.*, 1956, **4**, 11–22.

(Received April 29, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

VARIABLE FREQUENCY OF KNOWLEDGE OF RESULTS AND THE LEARNING OF A SIMPLE SKILL[1]

EDWARD A. BILODEAU AND INA McD. BILODEAU

Tulane University

Knowledge of results (KR) is known to be an important determinant of human skills learning and performance (**8**). Learning without KR has never yet been demonstrated, and rates and levels of learning have proved sensitive to its adequacy, variously defined. In a simple sense, there is an obvious likeness between KR and reward (punishment); each is the terminal effect of *S*'s behavior, dependent on *S*'s response, but also controlled by *E*. This likeness makes possible many analogous manipulations of time (**8**), consistency variables (**1, 2**), and frequency—as, to use the present study for illustration, varying the relative frequency of KR and varying the proportion of rewarded trials are procedurally analogous (**6**). In some cases, KR has effects similar to those of reward; for example, we find progressive improvement the greater the number of trials followed by KR (**1**), deterioration with its removal (**5**), and response shifts with arbitrary shifts in KR (**1**).

Despite many similarities between KR and reward, the two are not necessarily equivalent. There are differences as obvious as their likenesses. Reward is typically provided after one of a dichotomy of responses, KR more often varies with the *degree* of response error. Actually, some of the confusions arising out of comparisons of data from the two areas arise from the different types of tasks and procedures which have been used to date. In the human skills context, the task is usually one of learning to make graded responses by means of a graded error signal, KR being a quantitative index of how and by how much subsequent behavior should be modified. In the KR study, verbal instructions to *S* generally define the general problem, limit the response types, and establish the range within which the correct response lies. Absence of KR does not usually signify anything at all. On the other hand, in studies of reward, gradations of response are commonly irrelevant, a common reward being administered for any one of many responses meeting a broadly defined criterion such as, "turning right." Once we have a number of studies undertaken with comparable operations, there will be better opportunity to compare the operation of reward and KR at a theoretical level.

The purpose of this study was to compare the effects of absolute and relative frequency of KR. The task chosen for study was simple, requiring *S* to learn to move a lever a certain distance. The *S*s were given the magnitude and direction of the error on each trial or on a set proportion of trials. In the latter case, it is obvious that between KR trials *S* made responses at least qualitatively

[1] Preparation of the manuscript was supported by a grant of the University Council on Research of Tulane University. The data were collected while the authors were employed at the Air Force Personnel and Training Research Center. The present paper is based upon one reported before the Midwestern Psychological Association in 1956. A similar, but unpublished, study has been performed by E. I. Jones.

similar to those associated with KR trials. Whatever the KR schedule, KR informed *S* of his last error.

The hypothesis tested was that in a relatively simple, discrete motor-learning task, the effect of a KR upon the response tendency is immediate and without additional beneficial effect upon later non-KR responses. On the other hand, there is a list of potential negative considerations. With uncontrolled sources of external information, or with the growth of inhibitory (or extinction) effects, performance might be related to the relative frequency of KR. Further, interference from non-KR trials might reduce the effectiveness of the occasional KR; or the wider spacing of KR trials might reduce motivation. Nonetheless, since performance improves progressively, rather than stepwise with a succession of closely spaced KR's (**1, 2, 5**), learning can be expected from infrequent KR's.

Method

Subjects.—The *S*s were 273 basic trainee airmen at Lackland Air Force Base, all without previous experience with the apparatus.

Apparatus.—The lever apparatus has been fully described previously (**3**). In brief, it consists of a 14.5-in. lever suspended vertically, and shielded from *S*'s view. Moving the lever required a force of 20 lb. against the lever handle. One pull constituted a trial.

Design and procedure.—Three groups of 78 *S*s were used in the main investigation and 39 *S*s in a supplementary group were tested shortly thereafter. The KR was provided as follows for the groups of the main experiment: on all trials (Group 100), every third trial (Group 33), every fourth trial (Group 25). The KR was given to all groups on Trial 1 (after the first response) and thereafter KR followed the above schedule until 10, 30, and 40 trials had been administered to Groups 100, 33, and 25, respectively. All groups thus received 10 KR's. In analogy to animal-reward procedure, the present procedure represents administration of KR at a fixed ratio. The supplementary group, Group 10, was run immediately after data were collected for the other groups. Group 10 received KR for every tenth trial of its 100 trials, starting with the first KR on Trial 1.

For all groups the optimal lever travel was 33.57° of arc, but *S* was not informed of this—only that he was to find out how far to pull the lever to get a hit. On trials when KR was given, *E* gave *S* a verbal report of the magnitude of the error (rounded to the nearest whole number).[2] In addition, positive errors were read as "too high," negative as "too low." The task set *S* is thus one of minimizing the reported error.

The trial cycle was 20 sec., fractionated as follows: from ready signal to pull signal, 5 sec.; from pull signal to response completion, time as required by *S* (3–4 sec.); and from response completion to the next ready signal, time remaining within the 20 sec. When KR was given, it was interpolated 5 sec. after response completion.

Results and Discussion

The data reported are all in terms of degrees of arc. Of greatest interest are the data obtained on KR + 1 trials, i.e., trials immediately after those on which KR was given. These data provide a measure of the effect of KR when absolute frequency of KR is constant though relative frequency

TABLE 1

Mean Absolute Errors in Degrees of Arc

Trial KR + 1	Group*			
	100	30	25	10
1 st	9.01	10.01	8.71	10.88
2nd	6.65	6.18	6.11	5.87
3rd	3.99	4.47	4.86	3.83
4th	3.73	3.73	4.08	3.19
5th	3.33	3.51	3.87	2.32
6th	3.04	3.14	4.16	2.87
7th	2.63	2.75	3.11	3.28
8th	3.02	3.21	3.19	2.05
9th	2.21	2.83	2.83	2.64
10th	2.45	2.85	3.11	2.43

* For Group 10, $N = 39$; otherwise, $N = 78$.

[2] Actually, *S*'s error was not reported in degrees, but, by linear transformation, according to a scale graduated into 100 units. The conversion factor from errors in degrees of arc to error reported is 1.34.

is variable. Mean absolute errors in positioning the lever on the 10 KR + 1 trials are presented in Table 1. Leaving Group 10 out of consideration for the moment, it is obvious that there is little difference between groups on any trial. The data of Groups 100, 33, and 25 on KR + 1 trials were subjected to a Groups-by-Trials analysis of variance (Lindquist Type I [7, pp. 267–273]) in order to evaluate the effect of relative frequency of KR. The ratio of the mean squares of Groups and Subjects-within-Groups was 1.26, a value considerably below the .05 level with 2 and 231 *df*.

Group 10, run separately from the other groups, reverses, if anything, the slight trend in the main experiment: Group 10 generally has the smallest error on a given KR + 1 trial and has the smallest grand mean error over the ten trials. The Subjects-within-Groups error term from the analysis of variance mentioned above was used as the basis for setting up a confidence interval for the differences between means based on *N*'s of 78 and 39. Even

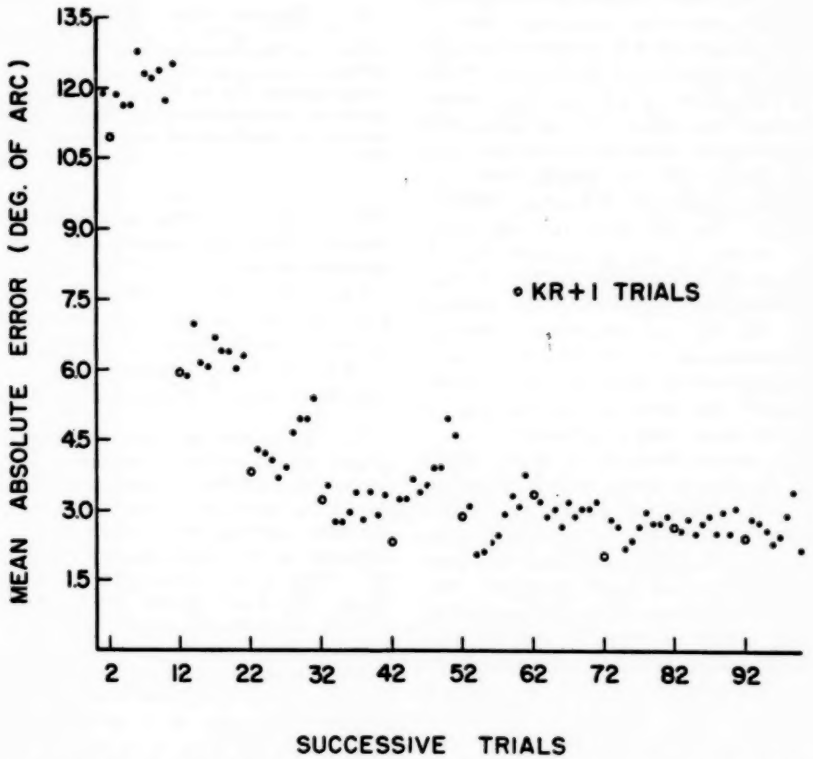


FIG. 1. The effects of KR trials (open circles) and non-KR trials (dots) are shown for the acquisition of a simple skill. The data are from Group 10, KR being administered after the first response and thereafter following every tenth response.

the greatest difference between grand KR + 1 means (Group 25 vs. Group 10) failed to exceed the .05 confidence interval.

In short, the findings do not support the idea that errors in pulling the lever varied as a function of variation in relative frequency of KR, absolute frequency constant. Examination of the algebraic errors on KR + 1 trials also failed to support any contention that behavior was affected by the partial KR schedules. The algebraic errors were susceptible to trial-by-trial variations in sign and did not seem to provide as stable and descriptive a measure as absolute deviations.

The errors on trials not directly associated with KR are also of interest and necessary to a description of the overall effect of KR on error reduction. The effect of omitting KR can best be gathered from the data of Group 10, the group with the longest series of responses between KR's. The 100 mean absolute errors of Group 10 are plotted in Fig. 1. There is no evidence in Fig. 1 that performance improves in the absence of KR for the preceding response. Performance, on the contrary, tends to deteriorate after KR + 1 trials, though the greater deterioration is in the early stage of practice.

A second illustration of the effect of practice without KR is given in Fig. 2 where the group means are plotted for Trials 2 through 11 (the Trial 1 response was blind and was followed by KR for all groups). The curves of Groups 25 and 33 are broken after a KR. Groups 25 and 33 show the step-like function attributable to the introduction of KR at selected points along the trial dimension. The step-functions of Groups 25 and 33 over Trials 2–11 are distinct from the trend of either Group 100 (KR after every response) or Group 10 (second KR introduced *after* the eleventh response).

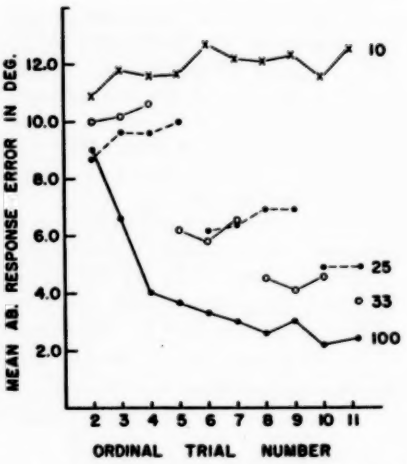


FIG. 2. The effect of relative frequency of KR upon a measure of response error. All groups received KR for the Trial 1 response, thereafter (except for Group 100) breaks in the curves represent the introduction of additional KR's.

Figure 2 clearly shows also that, with ordinal number of the trial constant and absolute frequency of KR variable, response error is negatively related to frequency of KR.

The major finding, that learning is related to the absolute frequency, and not to the relative frequency of KR, should be compared with partial-reinforcement findings with reward or UCS schedules as the experimental variable (6). With a few exceptions, such as Denny (4), *E* has related the response variable to successive trials, trials that included both reward and nonreward. Denny's analysis, which is comparable with the present one, gave similar results. For his study with the T maze, he reported that learning is equally good for schedules of 50% and 100% reward, provided performance is plotted against

rewarded trials (and provided there is no bias attributable to secondary reinforcement). In most studies in the reward area, the data are presented for blocks of trials, and it is impossible to distinguish between rewarded and unrewarded trials.

Summary

An experiment was performed to test for the effects of relative and absolute frequency of knowledge of results on learning a simple skill. The effects were tested by varying the proportion of trials on which KR was given, while keeping total number of KR's constant.

The analysis of results showed *learning* to be independent of relative frequency and positively related to absolute frequency. The former finding means that the learning effect of a KR is the same whatever the dispersion of KR, provided number of previous KR's is held constant. Inasmuch as the non-KR trials had no effect upon the responses of succeeding KR trials, it was not necessary to raise the issue of secondary KR during non-KR trials, interference from non-KR trials, nor differential motivation and inhibitory processes. It is emphasized that the findings were obtained from a fairly simple task situation without contamination of external sources of information beyond *E*'s control.

References

1. Bilodeau, E. A. Acquisition of two lever-positioning responses practiced over several periods of alternation. *J. exp. Psychol.*, 1953, **46,** 43–49.
2. Bilodeau, E. A. Motor performance as affected by magnitude and direction of error contained in knowledge of results. *J. Psychol.*, 1955, **40,** 103–113.
3. Bilodeau, E. A., & Ferguson, T. G. A device for presenting knowledge of results as a variable function of the magnitude of the response. *Amer. J. Psychol.*, 1953, **66,** 483–487.
4. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36,** 373–389.
5. Elwell, J. L., & Grindley, G. C. The effect of knowledge of results on learning and performance. *Brit. J. Psychol.*, 1938, **29,** 39–53.
6. Jenkins, W. O., & Stanley, J. C. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47,** 193–234.
7. Lindquist, E. J. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.
8. Lorge, I., & Thorndike, E. L. The influence of delay in the after-effect of a connection. *J. exp. Psychol.*, 1935, **18,** 186–194.

(Received May 7, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

INFORMATION PROCESSING AND PROBLEM SOLVING[1]

CHARLES M. SOLLEY AND FRED W. SNYDER

The Menninger Foundation

In many problem-solving tasks, complex visual discriminations are required. For example, in order to solve a jigsaw puzzle *S* must discriminate among the many pieces, randomly distributed in front of him, as to their form and size. The task is reduced to just this type of discrimination if the pieces are colored alike. The present studies represent attempts to study the relationships between this type of discrimination, time required to solve jigsaw puzzles, and the number of pieces composing such puzzles.

Experiment I

The purpose of this study was to provide basic information concerning the relationship between "discrimination time" and the number of alternative pieces in a puzzle. On the basis of Hyman's (**3**) studies it was suspected that the relationship would be logarithmic, but there were some fundamental differences between his procedures and the present ones so it was decided to let the empirical results determine further hypotheses. Obviously, in solving a jigsaw puzzle there is time required for discrimination and time required for movement. Therefore, the term "discrimination time" is used here reservedly.

Method

Eight adult, non-college-sophomore *S*s served in this study. The apparatus consisted of a 35-piece jigsaw puzzle painted a flat gray color to eliminate color patterns as a source of information. From this puzzle, five random samples of 2, 4, 8, 16, and 32 pieces each were drawn with the single restriction that each selected piece join with at least one other in the sample. The purpose of choosing five random samples was to approximate experimentally Brunswik's (**1**) "object sampling" of such problem-solving situations.

A given puzzle was randomly distributed in front of *S* and covered while *S* looked away. One piece of the puzzle, randomly chosen, was handed *S* with instructions that he fit it to one other piece from the puzzle. The remaining pieces were exposed and *S* worked until he had two pieces fitted together. The *E* recorded, from a stopwatch, the time required to the nearest second. Then another puzzle was presented and the process was repeated until all puzzles had been sampled.

Results

The average time required for "discriminating" is shown in Fig. 1 as a function of the number of pieces presented. Figure 1 shows that "discrimination time" (in minutes) is a *linear* function of the number of pieces to be discriminated among in this task. Fitted by the method of least squares these data yielded the equation:

$$t = .025n + .080.$$

Thus, the results of Exp. I indicate that "discrimination time" increases *linearly* with the number of pieces in a puzzle. The problem now arises as to how such data can be used to predict problem-solving behavior *within* a given puzzle. It can be argued that, when *S* has effectively made a single "discrimination" in a puzzle, he has reduced the number of pieces to be discriminated by one piece and continues making discriminations until all the pieces have been

[1] This research was partially supported by a research grant (M-715C) from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service.

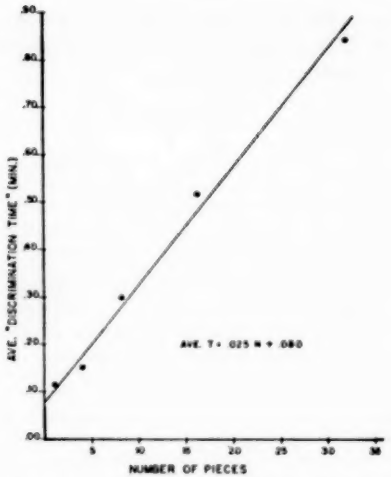


FIG. 1. Average "discrimination time" as linear function of number of alternative pieces to be "discriminated" among (least squares fit shown).

fitted together. The hypothesis for the next experiment was that cumulative time *within* a puzzle would act as if it were the simple additions of successive "discrimination time." Mathematically, this is derived as follows:

In Exp. I it was found that:

$$t = a(n) + b, \qquad (1)$$

where t is the time it took to discriminate among n pieces to fit two together. The constant b is the y intercept and is probably a measure of sheer work time to fit pieces together. The a is the rate of change in difficulty with increase in n.

The hypotheses for Exp. II and III state that:

$$\sum_{i=0}^{k} t_i = [a(n) + b] + [a(n-1) + b] + \cdots + [a(n-k+1) + b], \qquad (2)$$

where k is number of pieces fitted together; or equivalently, letting $\sum_{i=0}^{k} t_i = T_i$ or total cumulated work time:

$$T_i = a[(n) + (n-1) + \cdots + (n-k+1)] + bk \qquad (3)$$

$$T_i = a\left[kn - \frac{(k-1)(k)}{2}\right] + bk$$

$$T_i = ank - .5ak^2 + (.5a + b)k \qquad (4)$$

$$\therefore \quad \text{As } k \to n,\ T_i \to .5an^2 + (.5a + b)k + 0$$

which is a parabolic function of second degree.

$$\therefore \quad T_i \cong mk^h, \quad 0 < k \leq n,\ 0 < h \leq 2 \qquad (5)$$

which should predict the present data. Equation 4 could be used for an exact prediction, but it is easier to work with Equation 5.

EXPERIMENT II

Method

The hypothesis formulated on the basis of Exp. I was tested in this study. The materials consisted of the puzzles used in Exp. I. Figure 2 shows the theoretical curve predicted for each size puzzle on the basis of Exp. I. Five *S*s worked the puzzles and *E* recorded time ac-

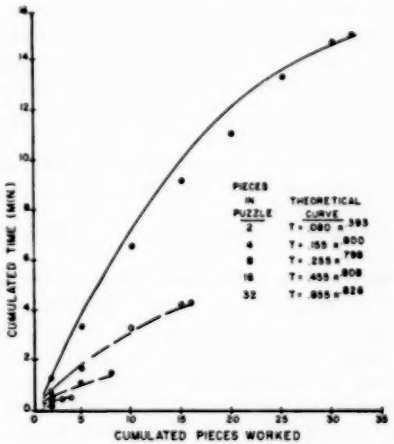


FIG. 2. Theoretical curves for cumulative working time, derived from successive integrations of area under curve in Fig. 1.

cumulated for 2, 5, 10, 15, 20, 25, 30, and 32 pieces, or until *S* solved the puzzle. Cumulative time was recorded for each piece of the 2- and 4-piece problems.

Results

The average cumulative curves are shown in Fig. 2. This figure shows that time accumulated while working on a given puzzle fits the theoretical curves generated on the basis of discrimination-time data found in Exp. I.

The results of this study indicate that *within* the process of solving a given puzzle successive "discrimination times" are additive, or operate *as if* they were additive. However, the results do not tell anything directly about how total "discrimination time" for solving a puzzle is related to relative difficulty *among* puzzles of varying number of pieces. Experiment III was undertaken to answer this question.

Experiment III

The problem facing the *E*s at this point was how to relate the results of Exp. I and II to differences between puzzles of various sizes. Fitting the curve defined by the final, theoretical cumulative "discrimination times" (based on Exp. II), i.e., Equation 5, and number of pieces, the following theoretical function was derived:

$$\text{Log}_2 T = 1.6664 \text{ Log}_2 n - 4.2960,$$

where T is total time taken to work a puzzle and n is the number of pieces in that puzzle. It should be noted that $\text{Log}_2 n$ is the number of bits of information in the puzzles, assuming all pieces equally probable of being used. It should be parenthetically added that this use of the term "information" seems legitimate, but the authors recognize an argument may be made against such usage (**4**). It is believed that the present usage is appropriate.

Method

Four *S*s were run, two of whom were the authors and two of whom were housewives who knew nothing of the purpose of the study.

The same samples of 2-, 4-, 8-, 16-, and 32-piece puzzles in Exp. I were used. In information theory terms these were 1-, 2-, 3-, 4-, and 5-bit problems. In addition, a 500-piece puzzle (nearly 9 bits) was used.

The total 25 puzzle samples were randomly presented, one at a time. The *S* was told that a jigsaw puzzle would be put in front of him and that he was to try to put it together as rapidly as he could. The *S* turned his head while *E* placed the mixed pieces of a given puzzle on the table in front of *S* and covered it with a large brown manila envelope. He was then instructed to turn around, uncover the puzzle, and begin. Time was scored from the moment *S* uncovered the pieces until he fitted the last piece into place. Solution time was recorded to the nearest second. When *S* reported feeling tired, he was allowed to rest before going on. This procedure was continued until all 25 puzzles had been worked.

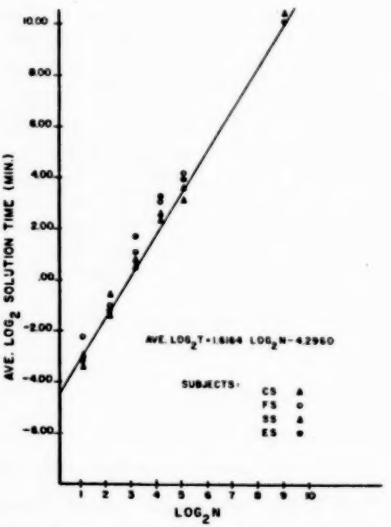


Fig. 3. Theoretical function, in linear form, between total solution time and number of pieces in various puzzles.

The two *Ss* who worked the 500-piece puzzle were presented with the pieces mixed together on a table and were told to begin. They stopped whenever they felt tired. Time was recorded when they began and when they stopped, and so on until completion of the puzzle.

Results

The solution times for each puzzle sample were converted to $\log_2$ and then were plotted as a function of the number of bits of information in the puzzles. A plot of the average log solution times against the number of bits put into the puzzles is shown in Fig. 3. Figure 3 shows a marked linear relationship between log solution time for puzzles and the number of bits of information (based on number of pieces) in the puzzles. It also shows that the empirical data fall closely about the theoretical curve. No statistical test of significance was computed since the evidence seems straightforward.

Discussion

It has been shown that discrimination time is a linear function of the number of pieces in a given jigsaw puzzle (Exp. I). It has also been shown that discrimination time operates *as if* it were additive *within* a puzzle (Exp. II). Finally, it has been shown that log solution time is a linear function of the number of bits of information among the pieces of puzzles (Exp. III) and that one can, theoretically, predict relative difficulty of various problem-solving tasks from a knowledge of how discrimination time functions in such situations.

It should be pointed out again that "discrimination time" may not be the best term to use. Clearly some aspect of complex form discrimination must be used to solve jigsaw puzzles, but there is considerable trial and error on a motorial level as well. However, it does appear that viewing jigsaw puzzles in this manner serves a useful function. The "bits" of information form an equal interval, dimensionless scale along which various puzzles can be arrayed. And simple properties of so-called discrimination time can be related to this scale of difficulty.

Why the typical logarithmic relationship was not found between "discrimination time" and number of alternatives, as reported by Hyman (**3**), needs some discussion. It is suggested that there is an enormous amount of "noise" in jigsaw puzzles. When the pieces to be discriminated among are randomly presented, they are disarrayed spatially with respect to the coordinates of the piece to be fitted. Though it has never been investigated as such, it could well be that spatial disarray can act as "noise" confusing the form discrimination process. As noise increases, discrimination time should become more and more a linear function of the number of alternatives.

Although the present studies do show a link between information theory and a type of problem solving, there are also a large number of other problem-solving situations where information theory cannot be so clearly applied, e.g., the pendulum problem of Duncker (**2**). It would be interesting and possibly fruitful to attempt to find such linkages with other types of problems to see how far such reasoning can carry us into the complex realm of problem-solving behavior.

References

1. Brunswik, E. *Systematic and representative design of psychological experiments.* Berkeley: Univer. of Calif. Press, 1949.
2. Duncker, K. (Trans. by Lynne S. Lees.) On problem-solving. *Psychol. Monogr.*, 1945, **58,** No. 5 (Whole No. 270).
3. Hyman, R. Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, 1953, **45,** 188–196.
4. Quastler, H. (Ed.) *Information theory in psychology.* Chicago: Free Press, 1956.

(Received May 15, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

DISTANCE CONSTANCY [1]

OLIN W. SMITH

Cornell University

Distance constancy has been discussed previously, but only as a by-product of fractionation judgments (1). One of the fractionation problems was to divide one distance into two equal lengths. The *S*s did this so well that training with correction did not materially decrease the pretraining errors of judgment. A second type of distance constancy problem requires *S* to judge a far small extent as longer or shorter than a standard, near extent. This judgment has not been studied. The purpose of this study is to obtain an estimate of the errors of this judgment.

METHOD

The experiment was conducted in a hall 360 ft. long and 7½ ft. wide. The standard stimulus was a white sheet of oilcloth, 54 in. wide and 10 ft. long. The sheet was centered in the hall with its far edge 25 ft. in front of *S*'s position. The vertical visual angle subtended to an eye 5.5 ft. above the floor was 7° 43′ 44″. The variable stimulus was a 25-ft. strip of the same oilcloth. It was rolled on a small roller the back edge of which was 125 ft. from *S*'s position. The distance of 125 ft. was selected because *S*s could not consistently judge a 25-ft. strip at 250 ft. as longer than the 10-ft. standard at 25 ft. At 175 ft. the judgment was still unreliable.

An approximation method was used for determining the value of *S*'s match. The variable stimulus was exposed to *S* in lengths which were multiples of 3 in. He was required to judge the far distance as longer or shorter than the standard distance. No judgments of equal were accepted. Initially, each *S* judged a series of random lengths of the variable distance. The *E* kept a log of the serial number of each judgment, the distance judged, and *S*'s response of "longer" or "shorter." Random lengths occasionally grouped around a restricted portion of the possible lengths of the variable. After the first six or seven judgments, if no strip longer than 16 ft. had been exposed, that distance was next judged. Similarly, if no distance shorter than 6 ft. had been exposed, *S* judged a 6-ft. length. These judgments were included to insure the opportunity for correct judgments of both "longer" and "shorter" early in the total series.

After the first mixed judgment by *S* (a long and short at the same distance setting of the variable) *E*'s task was to find the range of mixed judgments, the area of indeterminancy. For each end of the area *E* had to satisfy the following criterion. Three correct judgments had to be obtained for each of two settings of the variable, 3 in. different in length. These two settings had to be greater (or less) than any previous incorrect judgments. In the case of five correct judgments followed by an incorrect judgment a new series of correct judgments for adjacent values of the variable distance had to be obtained. Values of the variable throughout its entire length were mixed in serial order of judgment while the criterion was being obtained in order to prevent the subjective scale of *S* from being distorted.

The *S*'s score was the approximation of his mean match of equality which was the midpoint of his area of indeterminancy. This was the score which entered into all computations.

Each *S* was instructed to judge whether the far stimulus was longer or shorter than the standard, and was told that he could not make the judgment of equal. No *S* was corrected or informed of the correctness of his judgments. The *S*s were told that the variable would be either longer or shorter than the standard, that in no case would they be of equal length, and that at times it might be difficult to judge. In that case *S* should guess, if necessary. In any case, *S* should use all cues available for determining whether the far strip was longer or shorter than the near strip.

The S was told to stand at a marked observer's station while making a match. He could move

[1] This research was supported in part by Contracts AF33(038)-22804 and Nonr 401(14) between Cornell University and the Audio Visual Research Division, Human Resources Research Laboratory, Bolling Air Force Base, Washington, D. C., and Office of Naval Research, Department of Navy, Washington, D. C., respectively. Permission is granted to the United States Government for reproduction in whole or in part for any purpose.

his head and shoulders without restrictions. After each judgment he should walk into an adjacent part of the hall from which neither the standard nor the variable stimulus was visible. Upon a signal from *E*, he should return to the observer's station. The *S*s were 23 paid college students who had 20–30 vision or better, corrected or uncorrected.

Results

The means of the mean matches for the 23 *S*s was 11.51 ft., $SD = 1.26$ ft., $\sigma_m = .263$ ft. The mean of 11.51 ft. was significantly different from the length of the standard 10 ft., by two-tailed t test ($P < .01$, $t = 5.612$, $df = 22$). The vertical visual angle subtended by the mean match was 0° 15′ 20″.

Discussion

The data prove that approximately correct distance matches, revealing a constancy of distance, can be made of relatively short extents at different distances from *S*. The validity of these data are dependent upon the adequacy of the approximation method. The assumption is that the midpoint of the area of indeterminancy is a valid index of *S*'s mean judgment of equality. Individual records were examined and found not to be skewed. Consequently, the error of approximation should not be great. The method is time saving since *E* rarely needed 40 judgments to obtain his criteria for both ends of the area of indeterminancy.

The error of +1.51 ft. is equal to a vertical visual angle of 0° 2′ 10″, which is 13% of the mean match and .46% of the visual angle subtended by the standard. This error term has several possible sources. One is the elevation of the roll of oilcloth at the far end of the variable distance. This could have been a stimulus for additional length of the variable. A second source is the error of discrimination of two variable extents (at the far distance) as being different in length. This is assumed to be functional since the unreliability of the preliminary judgments of the variable at 250 ft. was probably due to the minute differences in angle subtended to the eye by the different lengths of the variable.

The study of distance constancy is a more direct approach to the study of discriminations and judgments of distance than is the size constancy method. The latter method provides no direct basis for estimating the error terms which need be known for predicting the errors of everyday judgments of distance which are normal to the normal person.

Summary

Twenty-three *S*s judged a strip of oilcloth with the far edge 125 ft. from *S*. Its length was varied as longer or shorter than a standard sheet 10 ft. long, the far edge of which was 25 ft. from *S*. An approximation method was used to determine *S*'s mean match of equality. The mean of the mean matches of all *S*s was 1.51 ft. greater than the standard. This difference was significant by t test ($P < .01$). The results prove that distance constancy matches of short lengths at different distances from *S* can be successfully made.

References

1. Purdy, J., & Gibson, E. J. Distance judgment by the method of fractionation. *J. exp. Psychol.*, 1955, **50**, 374–380.

(Received May 26, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

APPARENT BACKWARD ASSOCIATION: A SITUATIONAL EFFECT

LOWELL H. STORMS[1]

Institute of Psychiatry, University of London

Although backward association effects have been consistently found in experiments involving serial rote learning (**2, 3, 7, 12**), recall of paired-associate learning (**4, 5, 11, 16, 18**), associative clustering in recall (**6, 13**), and tachistoscopic recognition (**9, 13**), the appearance of the phenomenon in mediation is in doubt (**1, 10**) and its nature has not been determined. In particular, it is not understood whether backward association is, in part at least, a primary phenomenon independent of forward associative connections or whether it may be entirely derived from the operation of forward associations. Whether backward effects are in part primary or entirely secondary, the conditions which might bring about "backward" associations as artifacts or secondary phenomena are worthy of study, for it may prove that a series of artifacts can be discovered to account for all cases of backward association as derived phenomena.

Following upon the research of Bugelski and Scharlock (**1**), and Russell and Storms (**15**) who found that forward associations produced clear mediation effects, the writer conducted an experiment to investigate whether backward associations might play a mediating role in facilitating the learning of paired associates. An incidental result of the experiment suggested an hypothesis about a "recency" effect which might produce backward sequences as artifacts, and a second experiment was designed and executed to test that hypothesis.

[1] The research reported in this paper was performed at the University of Minnesota as part of the work for the author's Ph.D. dissertation and was supported by Office of Naval Research Contract No. N8onr-66216 as part of a series of studies in the area of verbal behavior.

Experiment I

Method

For the purposes of this paper, the occurrence of a C—B temporal sequence when only B—C sequences have previously been experienced shall be called a pure backward association. Unfamiliar materials, such as nonsense syllables, must be used in forming pure backward associations for experimental purposes. On the other hand, when C—B is a sequence of two words and it is known that B—C has a high frequency in word association norms while the frequency of C—B is nil, we shall speak of an inferred backward association. Only this inferred type will be dealt with here.

Subjects.—Sixteen female students from a beginning psychology course and ranging from 18 to 26 yr. of age served as *Ss*. The *Ss* were assigned to two groups on an arbitrary every-other-one basis.

Apparatus.—Words and nonsense syllables used in this experiment were presented at 2-sec. intervals (4 sec. between complete presentations of a list) by means of a Hull-type memory drum set on a desk with a cardboard screen blocking *S*'s view of *E*. The *Ss* sat in a comfortable swivel chair from which the memory drum opening was easily visible.

Lists.—Let A represent a nonsense syllable; B, a stimulus word from the Kent-Rosanoff list; and C, the most frequent response to B in the revised Kent-Rosanoff norms (**14**). Ten B—C pairs were chosen such that C—B associations were quite infrequent in the norms. The median B—C frequency was 32%. The median C—B frequency was 0% and the maximum, 2%. Ten nonsense syllables were selected from the Glaze list with 0% association value (**8**). Each nonsense syllable was assigned to a B—C pair. Five control words (X's), unassociated with the B's or the C's, were chosen from the Kent-Rosanoff list.

TABLE 1

NONSENSE SYLLABLES, B WORDS, C WORDS, AND CONTROL WORDS USED IN CONSTRUCTING EXPERIMENTAL LISTS

Nonsense Syllables	B Words	C Words	Control Words
ZIL*	EAGLE	BIRD$_2$	KING$_1$
LAJ	PRIEST	CHURCH$_1$	KING$_2$
GEX	JUSTICE	PEACE$_1$	HAND$_2$
YOV	MUTTON	LAMB$_1$	LONG$_2$
DAX	MEMORY	MIND$_2$	TABLE$_1$
CEF	STEM	FLOWER$_1$	TABLE$_2$
MYV	JOY	HAPPY$_1$	BUTTER$_2$
JID	TROUBLE	BAD$_2$	HAND$_1$
VUX	BATH	CLEAN$_2$	LONG$_1$
WUB	TOBACCO	SMOKE$_2$	BUTTER$_1$

Note:—Response words in the initial list presented to Group 1 are followed by a subscript 1; for Group 2, subscript 2.
* All materials were presented in capital letters.

Two forms of List 1 were then constructed. Form 1 consisted of five A—C pairs and five A—X pairs. Form 2 was derived by replacing the five C's of Form 1 with X's and the five X's with the C's unused in that form. These two forms were matched for number of syllables, number of letters, and Thorndike-Lorge frequency (**17**).

The Test List was constructed of 10 A—B pairs. The A's were paired with the B's which corresponded to the C's paired with them in List 1.

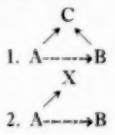
The nonsense syllables and the words paired with them for the two forms of List 1 and for the Test List are presented in Table 1.

Procedure.—The *S*s were instructed to pronounce each syllable and each word aloud as it appeared in the memory drum opening during the first presentation of a list, and on subsequent presentations to read the nonsense syllable and attempt to say the word following it before the word appeared.

Half the *S*s learned Form 1 of the first list, and half learned Form 2. Following this, a 4-min. rest period was spent in neutral conversation; then the Test List was learned.

The lists were presented to each *S* in three successive random orders before the first order reappeared. The repetitions continued until the criterion of three successive lists without error was reached on the initial list, one errorless presentation on the test list. The anticipation method, defining a correct response as saying a word before it appeared on the memory drum, was used. Errors were recorded. Since each *S* learned nonsense syllable-stimulus word pairs in the Test List, half of which had been preceded in the initial learning by Kent-Rosanoff response words paired with the same nonsense syllables and half preceded by learning unrelated words paired with the syllables, intra-individual comparisons were possible between unmediated (in the backward sense) pairs on the test list and pairs for which mediation was possible. In effect, each *S* served as her own control.

For each *S*, then, half the A—B's were learned after learning to associate A's with C's and half were learned without such prior training, the A's having been paired with irrelevant control words. For the Test List, the situation can be represented as follows, where solid lines indicate existing connections and dotted lines represent ones being learned:



Where A—C trials preceded the learning of A—B, the first diagram applies; for pairs where A—X trials preceded the learning of A—B, the second applies. The comparison of the number of correct responses while learning the A—B pairs in Situation 1 with the number for Situation 2 was the criterion measure for each *S*. The number of correct Situation 1 responses minus those for Situation 2 will be positive if facilitation occurs.

The purpose of this experiment was to determine whether effects of backward associations are observable in the mediation of paired-associate learning.

Results

Comparability of the two groups was checked in two ways: by testing the difference between their mean numbers of correct responses on pairs containing control words in the first list; and by testing the difference between their mean total correct responses, irrespective of mediation, on the Test List. Means and *SD*'s for the control words are presented in Table 2. The groups did not differ significantly ($P > .30$). Regarding the total correct responses on the Test List, Group 1 averaged 57.25 ($SD = 32.26$) and the Group 2 mean was 45.88 ($SD = 18.25$). This difference is also not significant. Although Group 2 averaged fewer

TABLE 2

MEDIATION INVOLVING BACKWARD ASSOCIATIVE LINKS: COMPARISON OF MEDIATED AND NON-MEDIATED PAIRS

Measure	No. Correct Anticipations			
	Mean	*SD*	*t*	*P*
Mediated minus non-mediated pairs	5.94	5.65	4.07	<.01
Control words				
Group I	78.25	26.51		
Group II	64.38	34.18		
Diff.	13.87	16.36	.85	N.S.

correct responses on both measures, it cannot be concluded that the two groups differed in learning facility.

The major results are summarized in Table 2. As indicated in the table, learning of the mediated pairs in the test list was clearly superior to the learning of unmediated pairs. The t of 4.07 is significant at the .01 level. The difference was most marked early in the learning trials. The percentage of correct responses in the first trial which were members of mediated pairs was 74.2, which is significantly different from 50% ($P < .01$). Figure 1 shows the number of correct responses for mediated and unmediated pairs as learning progresses.

All *S*s were interviewed after the experiment, and all but one revealed that they were conscious of some mediating connections which facilitated the learning of the test list. Asked to recall the word pairs which aided learning, the *S*s reproduced them significantly more often in the order of their appearance in the experiment (backward with respect to the norms) than in the B—C direction. Of the recalled pairs, 80.9% were recalled in the C—B direction ($P < .01$).

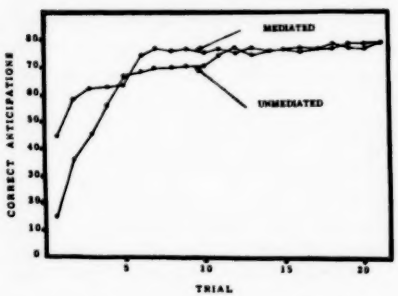


FIG. 1. Backward mediated facilitation of learning.

Backward associations inferred from word association norms have clearly been observed to act as mediators in the facilitation of paired-associate learning. The interview recall of more backward than forward associations, however, raises the question whether recent experience with members of associative pairs might raise response strengths temporarily and produce artificial backward associations. To investigate this possibility, a second experiment was performed.

EXPERIMENT II

Method

Two sheets of paper stapled together were distributed to each of the 59 (34 female and 25 male) students in a summer session beginning psychology class who volunteered to remain after a lecture for an experiment. On the first sheet were 14 words in one column, while the second sheet contained 36 words, each preceded by a number and followed by an underlined blank space, as on the forms used for obtaining association test data. The 14 words on the first sheet were B words, 10 of which were the same as the B words of Exp. I. Fourteen C words were printed on the second page along with 14 unrelated Kent-Rosanoff words as controls and 8 unrelated filler words. B words, C words, and control words are listed in Table 3.

The experiment was performed in three steps. (*a*) The *S*s, having been told as the sheets were distributed that they must not look at the second page until instructed to do so, were directed to follow along reading the words on the first page as the examiner read them aloud. (*b*) Instructions were given to tear off the first sheet and lay it upside down so

TABLE 3

B WORDS, C WORDS, AND CONTROL WORDS USED IN EXP. II

(B—C is a high frequency association while C—B is an improbable one)

B Words	C Words	Control Words
Mountain	Hill	Lion
Eagle	Bird	Sleep
Tobacco	Smoke	Blue
Bath	Clean	Foot
Hungry	Food	Moon
Scissors	Cut	Comfort
Butterfly	Moth	Hand
Mutton	Lamb	Cottage
Trouble	Bad	Health
Stem	Flower	River
Justice	Peace	Window
Priest	Church	Working
Joy	Happy	Baby
Memory	Mind	Head

the list could not be seen, then to write the first response that came to mind for each of the 36 words on the second sheet. (*c*) Then *Ss* were asked to turn over the sheet and write as many words as could be recalled from the first list. This step was included to give the appearance of a retroactive inhibition experiment; *Ss* had been told the experiment involved both association and recall.

Since normative data are available for the frequencies of responses to C words among beginning psychology students not having seen or heard the B words just previously, C—B frequencies could be compared between the experimental situation, i.e., just after reading B words, and the norms for a comparable group of *Ss*. As a check on the comparability of groups, frequencies of primary responses to the 14 Kent-Rosanoff words included in the 36-word experimental list were compared between another normative group and the present group of 59 people. These Kent-Rosanoff words with their most frequent responses will be called *control pairs*.

Results

Experimental and control pairs with associative frequencies for the two groups are presented in Table 3. Percentages of the responses to the C's that were the appropriate B's were computed for the normative data as well as for the present group. For purposes of the statistical tests, all percentages were converted by the arc sine transformation. Differences between the arc sines in the experimental and the normative data were obtained for each C—B pair, and a t test was computed from this distribution of differences. The mean difference was 8.88 (SD = 5.40). The t of 5.94 with 13 df is significant at the .01 level. That is, the C—B association was produced significantly more often under the experimental conditions, when B had been encountered just previously, than in the norms where it had not. Every one of the 14 C—B associations showed an increased frequency as compared with the norms, so even a crude sign test is highly significant. The mean frequency for the C—B associations in the norms was 1%, while the corresponding figure for this experiment was 19%.

In order to test whether the above striking results might be due to peculiarities in the sampling of students, the 14 control pairs were compared with respect to frequencies in the experiment and in the normative data. Although the normative group for these control pairs was not the same as that for the C—B pairs, it was quite similar in composition (students in beginning psychology classes at the same university) and much larger (N = 1008). The difference for the control pairs was in the opposite direction of the one observed for the experimental pairs, and the t of −2.01 was not significant ($P > .05$). This result does not suggest any vagaries of sampling which could account for the results. The median normative frequency of the control pairs was 21%, while their median frequency in the experiment was 17%.

A t test of the difference between the mean difference of 8.88 for the

experimental pairs and that of −1.88 for controls was significant ($t = 6.09$, $P < .001$).

One more question must be examined. Is the sensitization effect of recent reading of a word indiscriminate, making it a more likely response to all other words without favoritism, or is there some specificity of this effect to related words? When, for every pair, the maximum frequency of the B word in response to any word in the list except the appropriate C is subtracted from the C—B frequency, 13 of the 14 differences are positive, i.e., in favor of the appropriate stimuli. The mean difference is 8.98, $SD = 6.11$. The t of 5.30 is significant at the .001 level. Apparently, reading a B word led to increased backward associations rather than producing indiscriminate sensitization. Some degree of specificity is demonstrated.

Discussion

Experiment I showed that inferred backward associations among words may act as mediating links in the facilitation of paired-associate learning. Experiment II investigated one aspect of the nature of the inferred backward associations. Experience with the response members of associations of very low normative strengths (such as backward associations are by definition) apparently brings about, in some unknown manner, a striking but temporary augmentation of the effective strengths of association. Since this "recency" effect is situational and transitory it fits the criteria for a performance effect as distinguished from a manifestation of changes in habit strength in the Hullian sense. The effect appears to be selective, acting only upon habits of some non-zero strength, and therefore it cannot be used as the explanation for the initial formation of backward bonds when previously unassociated materials are used.

In Hullian terms, the recency effect might be formulated as an augmentation of reaction potential resulting from the occurrence of a response low in the habit-family hierarchy for some given stimulus. If that stimulus is presented to the organism soon after the emission of the response, that response will be more probable of elicitation than would ordinarily be the case. The mechanisms of the recency effect remain unexplored, but it is clearly shown that such an influence might produce apparent backward associations when words are involved. In the first test trial of the mediation experiment, *Ss* viewed and vocalized the B words, which might well have made the B words more likely of elicitation through A—C—B sequences in subsequent trials. Thus, what appear to be backward links in the mediation process are explicable as results of temporarily increased strengths of forward associations.

These considerations may also cast doubt on Russell and Storms' interpretation of their data (**15**). They set up inferred B—C—D associative sequences and showed that the learning of A—D was facilitated after learning A—B. Although B—D associations were of negligible frequency in the norms, recent experience with D might have served to make them more frequent in the experimental situation.

Summary

Two related experiments were conducted. To test the hypothesis that backward associations can act to facilitate the learning of paired associates, word pairs (B—C) with high associative frequencies in the B—C direction and very low C—B frequencies were employed in a memory drum experiment. Where A's are nonsense syllables, A—C and A—X (control word) pairs were first learned, then A—B pairs were learned in the test condition. If an A—C—B link contributes to learning A—B pairs, the A—B pairs for which A—C learning had previously occurred should be learned more rapidly than A—B pairs preceded by A—X learning. This expectation was confirmed at a high level of significance ($P < .001$).

The *Ss* recalled the facilitating associations significantly more frequently in the sequence of their being encountered in the experiment as

contrasted with the sequence of high associative strength in the norms. This led to the hypothesis that some sort of recency effect was contaminating the results. This hypothesis was checked in a second experiment, and it was concluded that the augmentation of associative strength by recent elicitation of the response members may produce apparent backward associations as artifacts.

References

1. Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44,** 334–338.
2. Cason, H. Specific serial learning; a study of backward association. *J. exp. Psychol.*, 1926, **9,** 195–227.
3. Garrett, H. E., & Hartman, G. W. An experiment on backward association in learning. *Amer. J. Psychol.*, 1926, **37,** 241–246.
4. Guthrie, E. R. Association as a function of time interval. *Psychol. Rev.*, 1933, **40,** 355–367.
5. Hermans, T. G. A study of the relative amounts of forward and backward association of verbal material. *J. exp. Psychol.*, 1936, **19,** 769–775.
6. Jenkins, J. J., & Russell, W. A. Associative clustering during recall. *J. abnorm. soc. Psychol.*, 1952, **47,** 818–821.
7. McGeoch, J. A. The direction and extent of intra-serial associations at recall. *Amer. J. Psychol.*, 1936, **48,** 221–245.
8. Melton, A. W. Materials for use in experimental studies of the learning and retention of verbal habits. Mimeo., Univer. of Missouri, 1940.
9. O'Neil, W. M. The effect of verbal association on tachistoscopic recognition. Office of Naval Research Contract No. N8onr-66216, Tech. Rep. No. 4, Univer. of Minnesota, 1953.
10. Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18,** 20–48.
11. Primoff, E. Backward and forward association as an organizing act in serial and paired associate learning. *J. Psychol.*, 1938, **5,** 375–395.
12. Raskin, E., & Cook, S. W. The strength and direction of associations formed in the learning of nonsense syllables. *J. exp. Psychol.*, 1937, **20,** 381–395.
13. Russell, W. A. Bi-directional effects in word association. Paper read at the SSRC conference on word association. April, 1955.
14. Russell, W. A., & Jenkins, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff word association test. Studies on the role of language in behavior. Office of Naval Research Contract No. N8onr-66216, Tech. Rep. No. 11, August, 1954.
15. Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49,** 287–293.
16. Storms, L. H. Backward association in verbally mediated learning. Unpublished doctor's dissertation, Univer. of Minnesota, 1956.
17. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers Coll., Columbia Univer., 1944.
18. Wohlgemuth, A. On memory and the direction of associations. *Brit. J. Psychol.*, 1913, **5,** 447–465.

(Received May 22, 1957)

Journal of Experimental Psychology
Vol. 55, No. 4, 1958

ERRATUM

In the article "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries," by A. M. Liberman, K. S. Harris, H. S. Hoffman, and B. C. Griffith, in the November, 1957, issue of this JOURNAL (*J. exp. Psychol.*, **54**, 358–368), a typographical error was made in each of the last three equations in the right-hand column of page 363. These equations should read:

$$P\ Corr_{(ABA)} = .5 + \frac{p_b^2 + p_d^2 + p_g^2 - p_b p_b' - p_d p_d' - p_g p_g'}{2}.$$

$$P\ Corr_{(ABB)} = .5 + \frac{p_b'^2 + p_d'^2 + p_g'^2 - p_b p_b' - p_d p_d' - p_g p_g'}{2}.$$

$$\bar{P}\ Corr = .5 + \frac{(p_b - p_b')^2 + (p_d - p_d')^2 + (p_g - p_g')^2}{4}.$$

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland

PROF. DR. PH. LERSCH, München; PROF. DR. F. SANDER, Bonn

Herausgeber für Österreich	Herausgeber für die Schweiz	Redaktion
PROF. DR. H. ROHRACHER, Wien	PROF. DR. R. MEILI, Bern	DR. C. J. HOGREFE, Göttingen

Band IV/3

JOHANNES C. BRENGELMANN, **Extraversion, neurotische Tendenz und Rigidität im Umkehrversuch (Prismenbrille)**

KLAUS CONRAD, **"Occulte" Phänomene im Lichte gestaltpsychologischer Forschung**

HEINRICH DÜKER UND REINHARD TAUSCH, **Über die Wirkung der Veranschaulichung von Unterrichtsstoffen auf das Behalten**

SIEGFRIED HADDENBROCK, **Zur Struktur und Dynamik der humanen (leib-seelischgeistigen) Gesamtsituation**

PHILIPP LERSCH, **Zur Theorie des mimischen Ausdrucks**

HEMMO MÜLLER-SUUR, **Psychopathologische Voraussetzungen zur Erfassung der organischen Komponenten schizophrener Psychosen**

HUBERT ROHRACHER, **Zur Einteilung und Systematik der psychischen Phänomene**

HANS THOMAE, **Längsschnittuntersuchungen zum Problem der Beziehungen zwischen körperlicher und seelischer Entwicklung**

UDO UNDEUTSCH, **Das Motivationsgeschehen bei schuldhaft normwidrigem Verhalten**

ALOYS WENZL, **Zur Psychologie des Geistes unserer Zeit**

KURT WILDE, **Über die Mitnahme im binokularen Tiefensehen**

KONRAD ZUCKER, **Bannung und Gespenstererleben**

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.
Subscription is DM 64.—per volume.

VERLAG FÜR PSYCHOLOGIE • DR. C. J. HOGREFE • GÖTTINGEN
(Brentanoweg 10, Germany)

ERGONOMICS

HUMAN FACTORS IN WORK, MACHINE CONTROL AND EQUIPMENT DESIGN

A Taylor and Francis International Journal

The Official Publication of the Ergonomics Research Society

General Editor

A. T. WELFORD

Editorial Board

Belgium: Dr. H. BASTENIER. *France:* Professor R. BONNARDEL and Professor BERNARD METZ. *Germany:* Professor E. A. MÜLLER. *Great Britain:* M. G. BENNETT, Dr. W. F. FLOYD, Dr. W. E. HICK, Emeritus Professor Sir CHARLES LOVATT EVANS, and Dr. L. G. NORMAN. *Netherlands:* Dr. F. H. BONJER. *Sweden:* Professor S. P. M. FORSSMAN. *Switzerland:* Professor E. GRANDJEAN. *U.S.A.:* Professor H. S. BELDING and Professor P. M. FITTS.

Assistant Editor

Miss H. M. CLAY

Volume 1 February 1958 Number 2

Contents

The Strength of the Lifting Action in Man.....By R. J. WHITNEY

Physiological Methods for Estimating the Physical Work Capacity in Workers Especially of the Older Age Groups.....By IRMA ÅSTRAND

Use of Competitive Tests as a Method of Performance Research. By M. J. KARVONEN

Research on Human Factors in Road Transport.....By G. GRIME

Headlight Design.....By R. L. MOORE

Simplifying the Operator's Task as a Controller.....By A. W. BAILEY

The Relationship Between Dial Size, Reading Distance and Reading Accuracy.....By K. F. H. MURRELL, W. D. LAURIE and C. MCCARTHY

Price £1 5s. 0d. per part

Subscription price per volume £4 15s. 0d. post free, payable in advance

Printed and Published by

TAYLOR & FRANCIS Ltd.

Red Lion Court, Fleet Street, London, E.C.4.

Agents for U.S.A. and Canada

Academic Press Inc., 111 Fifth Avenue, New York 3, N.Y., U.S.A.

$3.50 per part. $13.50 per volume.